

SIMULATIONS OF MIXED USE RAIL CORRIDORS:  
HOW INFRASTRUCTURE AFFECTS INTERACTIONS AMONG TRAIN TYPES

BY

SAMUEL L. SOGIN

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Civil Engineering  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Advisers:

Professor Christopher P.L. Barkan  
Senior Research Engineer C. Tyler Dick, P.E.

## **ABSTRACT**

Freight traffic on North American railroads has undergone dramatic growth over the past several decades and this trend is projected to continue for the foreseeable future. Meanwhile, there is widespread interest in use of existing freight railway infrastructure to provide higher-speed passenger rail service. The growth in these two types of traffic, both alone, and particularly in combination, is increasingly straining the capacity of the rail network. In many locations, new railway infrastructure construction will expand capacity to accommodate this growth. In others, new efficiencies may be gained through advances in operational practices and investment in new traffic control technologies that can extract additional capacity from existing infrastructure. Railroads are highly capital intensive so careful consideration must be given to maximizing the benefit from these investment decisions.

Defining railway line capacity is non-trivial. Characteristics of rail infrastructure such as the number of tracks, spacing and length of passing sidings on single track, the location of crossovers in double track, the traffic control system and other factors all affect the number, speed and performance of trains on a rail line. These interact with individual train operating characteristics and heterogeneity among train types to have major effects on throughput and service quality. These interactive effects can be quite complex and there is no fundamental understanding of many important relationships. Rail Traffic Controller (RTC), a railroad infrastructure and dispatching simulation software was used to conduct an extensive series of experiments investigating interrelationships between infrastructure configuration, traffic volume and mix, train speeds and delay. Design of experiments and non-linear regression statistical tools were used to maximize the information gained from the simulated data. Potential delay-based methods were investigated, as well as key highway capacity relationships that have railroad

corollaries. A framework from failure analysis was adapted to evaluate train delay distributions. A number of factors have the potential to affect railway performance; this research identified the most important affecting railway capacity, train delay and reliability. It presents investigations and results that enable comparison of their relative importance and effects thereby enhancing our understanding of several fundamental relationships between infrastructure, train characteristics and operations that affect railway performance.

*“No other improvement that reason will justify us in hoping for,  
can equal in utility the rail road.”*

-Abraham Lincoln, March 9<sup>th</sup>, 1832

## **ACKNOWLEDGEMENTS**

Most importantly, I thank Dr. Christopher P.L. Barkan for continuing to develop and grow railroad related research and classroom learning at the university level in North America. Dr. Barkan strives for excellence in his graduate students and I was very honored when he asked a confused economics major to pursue graduate studies in Civil Engineering. Dr. Barkan provided me with freedom to conduct exploratory research in railway capacity and constantly brainstorm new ideas worth investigation. He has been instrumental in further developing my writing communication, and presentation skills. I hope that this thesis will stand tall on his bookshelf next to the other great theses and dissertations of those who came before me.

C. Tyler Dick joined RailTEC during my second year. Tyler has a unique ability to absorb new ideas, methodologies, and analysis tools. He often challenges and questions my most current analysis and through these critiques, my work was advanced forward. J. Riley Edwards was the first person at the University of Illinois to suggest that I study railroad engineering when I was still an economics major. He encouraged me to continue attending the William W. Hay Seminars and eventually helped me gain the opportunity to work as an undergraduate assistant to Mark Dingler in his M.S. thesis research on railway capacity analysis.

Mark Dingler of CSX Transportation has given me tremendous technical insight. My work has been heavily influenced by Mark's initial work on railroad capacity at the University of Illinois and his feedback as a practitioner in the field. In turn, Mark served as a research assistant to Dr. Yung-Cheng Lai during his Ph.D. studies at UIUC and I consider Dr. Lai my research grandfather. Yung-Cheng is now an Associate Professor at National Taiwan University has been a great co-author on many of my papers assisting me in the technical development of this work. Additionally, he hosted me in Taipei, summer 2011.

Dr. David Sogin, my father, was always a willing listener to my latest research issue. Besides being a captive listener, his experience in conducting industrial laboratory experiments has influenced my research. He suggested using design of experiments and regression statistical techniques to determine key effects with less data.

Brennan Cauhgron, Samantha Chadwick, and I were undergraduate students together at the University of Illinois. We have supported each other in and outside our academic interests. We were joined by Amogh Shurpali, Chris Rapp, and Brandon Van Dyk in graduate school. Working with this cohort has been a privilege and honor, and I wish them the best in their future endeavors. Ivan Atanassov and Scott Schmidt were able to find time from their undergraduate studies to assist me in various ways. They created schedules, learned simulation tools, and conducted their own experiments. The students in B118 are the best.

Financial support for my research came from several sources. I was initially supported by a CN Graduate Research Fellowship in Railroad Engineering. Subsequently I was supported by my work on a project investigating technical challenges and research needs related to shared corridors for passenger and freight operations funded by the Federal Railroad Administration, Office of Research and Development, Broad Agency Announcement program. Later in my studies, I received support from the National University Rail (NURail) Center, a US DOT RITA University Transportation Center. I am also grateful to John Gray of the Association of American Railroads for his support of research on railroad capacity analysis at UIUC RailTEC. Finally, I thank Eric Wilson and others of Berkeley Simulation Software for providing an academic license to UIUC RailTEC for use of Rail Traffic Controller (RTC). This software has been instrumental in conducting my series of experiments, and previously helping to launch the rail capacity research program at RailTEC.

# TABLE OF CONTENTS

<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
1.1 Objectives and Scope .....	2
1.2 Organization.....	2
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>7</b>
2.1 Capacity Models.....	7
2.2 Scheduling.....	11
2.3 Delays Due to Heterogeneity .....	15
2.4 Influence of Infrastructure .....	20
<b>CHAPTER 3: RAILROAD CAPACITY METRICS.....</b>	<b>23</b>
3.1 Introduction .....	23
3.2 Differences between Highways & Railways .....	24
3.3 Types of Capacity .....	25
3.4 Traffic Flow.....	27
3.5 Delay Relationships .....	31
3.6 Transportation Economics .....	37
3.7 Performance Metrics .....	40
3.8 Spatial Analysis.....	44
3.9 Simulation Case Study.....	46
3.10 Conclusion .....	51
<b>CHAPTER 4: RAILROAD SIMULATION TECHNIQUES FOR ACADEMIC RESEARCH .....</b>	<b>52</b>
4.1 Introduction .....	52
4.2 Inputs to RTC.....	52
4.3 University of Illinois at Urbana-Champaign RTC Standards .....	54
4.4 RTC Signals .....	58
4.5 .INPUT Files and .OPTION Files.....	59
4.6 Data Generation.....	61
4.7 Design of Experiments .....	64
4.8 Data Analysis .....	64
<b>CHAPTER 5: COMPARISON OF THE CAPACITY OF SINGLE AND DOUBLE-TRACK RAIL LINES USING SIMULATION ANALYSES .....</b>	<b>77</b>
5.1 Introduction .....	77
5.2 Methodology.....	78

5.3	Analysis .....	82
5.4	Correlation Analysis .....	91
5.5	Conclusion .....	94
<b>CHAPTER 6: ANALYZING THE TRANSITION FROM DOUBLE TO SINGLE TRACK WITH DESIGN OF EXPERIMENTS TECHNIQUES .....</b>		<b>96</b>
6.1	Introduction .....	96
6.2	Freight Model.....	107
6.3	Passenger Model.....	117
6.4	Combined Model.....	120
6.5	Future Work .....	121
6.6	Conclusion .....	123
<b>CHAPTER 7: ANALYZING THE TRANSITION FROM SINGLE TO DOUBLE TRACK WITH NON-LINEAR REGRESSION ANALYSIS .....</b>		<b>124</b>
7.1	Introduction .....	124
7.2	Simulation Methodology.....	131
7.3	Developing the Response Surface Model .....	137
7.4	Comparing Different Operating Conditions .....	144
7.5	Future Work .....	151
7.6	Conclusion .....	151
<b>CHAPTER 8: RELIABILITY APPROACH TO VARIATION IN TRAIN DELAYS .....</b>		<b>153</b>
8.1	Analogies from Failure Analysis .....	154
8.2	Distributions by Network Configuration .....	162
8.3	Modeling the Delay Distribution Transition.....	163
8.4	General Findings & Future Work.....	166
<b>CHAPTER 9: GENERAL FINDINGS AND FUTURE RESEARCH OPPORTUNITIES .....</b>		<b>168</b>
9.1	Future Research .....	168
9.2	Conclusion .....	176
<b>REFERENCES CITED .....</b>		<b>179</b>



# **CHAPTER 1**

## **INTRODUCTION**

Railroads provide low-cost, long-distance transportation services to numerous sectors of the North American economy. Companies are able to leverage this efficient transportation mode to improve their access to both domestic and international markets, enabling them to effectively compete in the global marketplace. Long-term freight transportation is projected to increase 80% by 2035 (AASHTO 2007). Throughout much of its history, coal was the largest source of traffic carried by railroads in the United States in terms of both tonnage and revenue. However, new trends in the economy have reduced the amount of coal used for power generation and corresponding transport needs. Meanwhile, intermodal and new traffic in alcohol and petroleum crude oil has undergone dramatic growth over the past decade (AAR 2013). These changes in freight flows compounded with a general increase in long-term demand for rail freight transport are straining the capacity of the railroad network.

Meanwhile, new passenger rail initiatives are proposing expanded use of the existing freight railroad network in order to provide new commuter, regional, and intercity service. As evidence of the broad interest, 34 states requested \$57 billion for passenger rail projects in 2009 ranging from conventional operations (79 mph) to new 220-mph high speed operations. Since then, over \$10.1 billion of federal funds have been appropriated to leverage state investments to develop passenger rail. \$6.2 billion has been allocated to planning studies for new routes, Amtrak's Northeast Corridor and improvements to existing passenger corridors. Most of these projects support expanded or upgraded regional, intercity, and commuter trains that share infrastructure and right of way with heavy tonnage freight trains (Federal Railroad Administration 2013a). This new growth in both passenger and freight traffic is increasing the

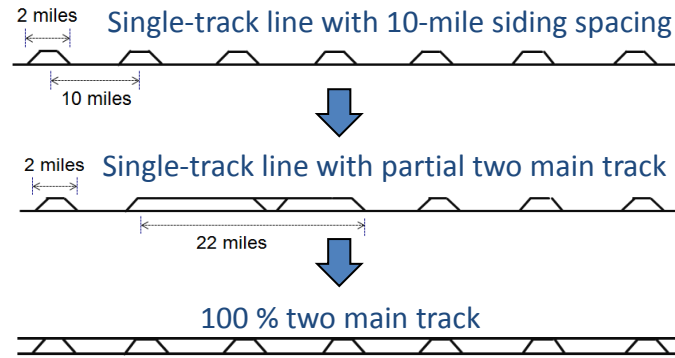
need for additional railroad capacity. However, simultaneous operation of freight and passenger trains on the same trackage can cause additional challenges (Saat & Barkan 2013) and disproportionately reduces capacity.

## **1.1 Objectives and Scope**

The objective of the research described in this thesis was to better understand certain fundamental relationships between railroad infrastructure, operations, and performance. A subset of the critical factors was investigated by conducting a series of systematic simulation experiments. A major focus of these studies was to determine the effects of higher passenger train speeds on freight traffic in various infrastructure configurations and how infrastructure changes affected capacity.

## **1.2 Organization**

This thesis is divided into two sections. Chapters 2 through 4 provide an overview of railway capacity relationships and metrics. This includes a comprehensive literature review and a guide to using rail traffic simulation software for academic research. Chapters 5 through 8 investigate effects of higher-speed passenger trains on freight traffic and how the infrastructure changes this impact. Chapter 5 introduces the topic by comparing single and double-track networks. Chapters 6 and 7 investigate upgrading a single-track line by systematically adding segments of two-main-track. These intermediate configurations become “hybrid” track configurations (Figure 1-1). Chapter 8 further investigates the transition from single to double track by analyzing the reliability of train arrivals with concepts borrowed from failure analysis.



**Figure 1-1: Progression from single to double track.**

### 1.2.1 Chapter 2

Sharing existing tracks between passenger and freight trains will increase train delays and have the immediate consequence of consuming the available capacity of that corridor without mitigation measures. This increase in capacity utilization can lead to scheduling issues, increased train delay, and smaller time windows to maintain infrastructure. Different traffic types can have different characteristics in terms of acceleration and braking performance, top speed, priority and on-time performance sensitivity. These unique characteristics place different demands on the freight infrastructure. This chapter includes a literature review and discussion of capacity planning, train scheduling, and train delays.

### 1.2.2 Chapter 3

Measuring the capacity of rail lines is complicated by the interrelationships between asset utilization, reliability, and throughput. There is not a single metric that captures intricacies of these relationships. Capacity can be evaluated by delay-volume relationships, utility models, or economic studies. For many case studies, railroads use parametric and simulation modeling to determine train delay per 100-train-miles and use this value to assess system capacity. However, this metric does not tell the full story, especially when comparing different train types. Highway

transportation planners use a different portfolio of metrics that can complement those used by railroad capacity planners. This chapter discusses the relationships among different railroad capacity metrics and considers some of the relative relationships with highway capacity metrics. These concepts are illustrated by simulating different train types operating on a representative freight line. Preliminary material from this chapter was presented in a paper at the 2012 Joint Rail Conference in Philadelphia.

### **1.2.3 Chapter 4**

This chapter presents a general methodology for how the railroad simulation experiments using Rail Traffic Controller (RTC) in the subsequent chapters were conducted. Additionally, this chapter serves as a guide on how to use the RTC software for academic research including descriptions of how to efficiently generate and analyze simulation data. A brief overview is provided on design of experiments and use of regression techniques to analyze train delay.

### **1.2.4 Chapter 5**

This chapter investigates the influence of track configuration – single versus double track – affects between passenger and freight trains interaction. The maximum speed of passenger trains has little effect on the performance of freight trains on single-track lines. However, in double-track configurations, the speed of the passenger train will have a major impact on freight train delays. Single track can show an asymmetrical delay distribution centered on an average run time, with very few trains arriving close to the minimum run time. A double track configuration can result in a delay distribution shaped similarly to an exponential distribution with many trains maintaining their minimum run times. In both single and double track, a higher passenger train maximum speed can lead to a greater range of possible travel times. These analyses can help

further understand the interactions between passenger and freight trains for current and future shared corridor operations. The material from this chapter has been accepted for publication in the Transportation Research Record.

### **1.2.5 Chapter 6**

A substantial portion of the routes in the United States rail network is single track with passing sidings. As traffic continues to grow, it will be necessary to add trackage in order to maintain network fluidity. In many cases, this will involve adding or extending sections of second mainline track. However, the investment required may not be justifiable or practical all at once; therefore the second track can be phased in over time by creating stretches of double track separated by single-track segments (Figure 1-1). Depending upon the traffic, the operational characteristics will transition from those similar to single track to those similar to double track. A response surface model for delay was developed to investigate this transition. Experiments were conducted to understand the effects of various factors including the amount of second main track, traffic volume, traffic composition, and the speed differential between train types. Design of experiments software (JMP) was used in conjunction with railway simulation software (Rail Traffic Controller 2012) to conduct the analyses. Most of the double-track performance characteristics were not achieved until nearly the entire second mainline track was installed. The results suggested a linear relationship between miles of second mainline track added and reduction in train delay. The influence nature of the various factors on the delay response surface provide insight regarding the nature interactions between passenger and freight trains on mainline trackage. Initial findings of this work were first presented at the 2012 Annual INFORMS Conference and an expanded presentation was given at the 2013 Joint Rail Conference.

### **1.2.6 Chapter 7**

This chapter also investigates the transition from single to double track; however, instead of a partial-factorial experimental design, a full-factorial design in conjunction with non-linear regression was used. This analysis relates the capacity benefit of double tracking to typical exponential delay-volume relationships. Adding sections of double track reduces train delay linearly under constant volume. Because of this linear relationship in delay, capacity will have a convex transition function from single to double track. Speed differentials between train types will have greater effects on capacity when the route is closer to full double track. Additionally, a substantial amount of infrastructure is needed to create the capacity for these higher speed trains as well as mitigate freight train delays. The material in this chapter was published in the conference proceedings for “The 5th International Seminar on Railway Operations Modelling and Analysis – RailCopenhagen”.

### **1.2.7 Chapter 8**

This chapter investigates the statistical distribution of train delay under various circumstances and considers the relationship of similar distributions used in failure and survival analysis. Key equations from these fields are adopted for analyzing train delays. The cumulative failure curve corresponds to the cumulative train arrivals. Hazard functions now measure the instantaneous conditional probability of an en route train arrival. Single track and double track configurations have different delay distributions as described in Chapter 5. A logistics model was developed to categorize delay distributions based on the amount of double track and traffic volume.

### **1.2.8 Chapter 9**

General findings and future research opportunities are summarized in this chapter.

## **CHAPTER 2**

### **LITERATURE REVIEW**

Rail capacity is often described as the maximum number of trains that can travel on a link in a network per unit time. This chapter explores previous research on rail capacity and train delays. The first section identifies the types of models used for railway capacity analysis. The second section introduces scheduling and its relationship to capacity. The third section highlights research on train delays due to heterogeneity in train types. The last section focuses on the relationship between infrastructure and capacity.

#### **2.1 Capacity Models**

Railway line capacity is often limited by some bottleneck along the route. In some segments, there may be a single bottleneck that limits capacity utilization for the remainder of the route. In other circumstances, the bottleneck may not be easily identified (McClellan 2007). Numerous approaches and tools have been developed to determine rail line capacity. Each of them has its strengths and weaknesses and has been generally designed for a specific application. Railway capacity tools can be categorized into three groups, (i) theoretical, (ii) parametric, and (iii) simulation. The precision of these models varies depending on the amount of data inputs and details incorporated. Some models may exactly emulate railway operations, while others may provide approximate answers quickly that can guide the planning process (Assad 1980). A simpler model may arrive at conclusions faster but lack the accuracy or precision necessary to make a business decision. Complex models with many inputs can take more time to develop and may provide accurate information.

### **2.1.1 Theoretical**

Theoretical models can be useful for determining initial estimates of railway line capacity in the beginning stages of the planning process as well as analyzing an entire railway network. Theoretical models are useful for analyzing passenger operations where most train movements are carefully planned and executed by timetable and less by a dispatcher. Most theoretical models measure the flow rate relating to the maximum amount of trains that can pass a bottleneck per unit time. In a single-track network capacity is often governed by the number and location of sidings. Early capacity models assumed the capacity of the railway line to be the traffic level when all sidings are occupied by stopped trains waiting for the mainline to clear (AREA 1921). Poole (1962) approximated the meet interactions at sidings to create a proxy for headway. He developed a simple formula that accounted for the time required for pairs of trains in opposing directions to alternate their use of the single track section. This model becomes more inaccurate if traffic demand is non-uniform or there is a directional bias to the traffic that changes over time.

The constraint of the “bottleneck” could be the signal spacing, where a blocking-time model could be used. The blocking-time model analyzes the route setup time and signal block occupation time that a train uses to traverse its route (Pachl & White 2004). These blocks are plotted on a time-distance plot and form a staircase pattern. Either a computer algorithm, or individual then compresses these staircase patterns in order to determine minimum train spacing. Shorter headways between trains improves capacity but increase the propagation of delay as trains are more likely to agglomerate (Carey 1994). Many schedule compression models only analyze one direction and one track at a time. Additionally, these models often assume that no trains are delayed. Mathematical models can also be used to estimate delay-minutes between



train types, (Chen & Harker 1990, Galaverna & Sciutto 1999, Huisman & Boucherie 2001). Theoretical models are faster and easier to communicate than delay-based-models; however, they often over simplify key stochastic processes of railway operations. Parametric models may address some of these disadvantages.

### **2.1.2 Parametric**

Parametric models use field or simulated data to generate a statistical model that can predict rail line capacity more quickly than full simulation. Generally, parametric models will handle more inputs than theoretical models. Regression analysis of train delays has been used to quantify effects on train delay for various operational factors. Prokopy and Rubin (1975) used single track simulation results to develop a multivariate regression model. Kruger (1999) used a similar approach with an updated simulation model and also summarized the data through multivariate regression. Both of these models were developed by varying only a single parameter at a time. Mitra et al. (2010) developed an 8-variable regression model using simulation results for single-track lines, but their model did not consider interaction effects between variables. They used a partial factorial experimental design in which siding spacing, siding uniformity, speed limit, traffic composition, and traffic volumes were varied (Murali et al. 2010). Lai and Huang (2012b) used regression and neural networks to model Rail Traffic Controller (RTC) simulation results from both a single and double-track network using a full-factorial experimental design that analyzed five factors at three different levels.

Parametric models can incorporate more variables than theoretical models and can be used to represent a greater portfolio of infrastructure configurations. However, parametric models may only describe average train delays or a single estimate of useable capacity. These

models do not provide accurate information on the performance of individual trains or groups of train types. A simulation model can provide more detailed information of individual trains.

### **2.1.3 Simulation**

Simulation modeling can incorporate many variables related to track, signaling, trains, and schedules and realistically emulate rail operations. These models can accurately predict train performance on mainlines and in terminals. They usually incorporate a train performance calculator that can estimate train acceleration and braking curves thereby enabling fairly accurate prediction of train movement along a line. Another module then resolves conflicts between different trains. Simulation software can be specific to railroad applications such as RailSys (Germany) and Rail Traffic Controller (United States). Alternatively, traditional resource modeling can be adapted to mimic railway operations. In this case, the railroad is approximated by batches, processes, resources, and queues (Lewellen & Tumay 1998, Kulick 2011). Other models have used a hybrid approach and with stochastic train arrival events into a network flow problem, and developed a global dispatching algorithm to route the trains. These models are computationally fast; however, they may not include the signaling system or realistic train acceleration performance. (Petersen 1982, Lu et al. 2004). Typical simulation outputs include run-time, delay, instantaneous velocity, and fuel consumption. Some simulations require strict train schedules prior to the simulation while others are capable of orchestrating train movements without a schedule and will delay trains as necessary. Certain simulations packages can estimate theoretical capacity by using algorithms to compress train schedules into a shorter time period and extrapolate the time saved to create more train slots. The capacity is then calculated as the maximum number of feasible train slots given the initial train schedule (Pouryousef & White 2013).

#### **2.1.4 Optimization**

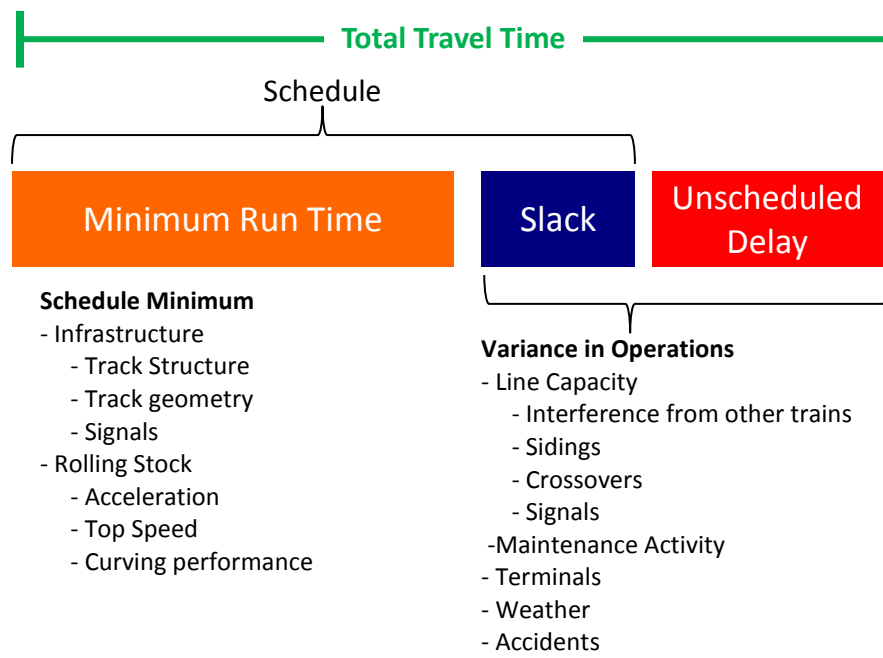
Optimization can be used as a supplement to the models mentioned above. These algorithms search for minima or maxima under a series of constraints. Some applications include designing dispatching algorithms capable of minimizing delays through a network. The departure times can be fixed (Higgins et al. 1996) or the schedule can be a decision variable (Burdett & Kozan 2006). These optimization models can include traffic mixture, station stops, and train lengths. Ekman (2004) developed capacity estimates by using discrete event calculation of train paths. Additionally, there are optimization techniques that can aid in planning a corridor. Once the cost and benefits of a set of projects are known, a selection model can choose projects to undertake based on a series of constraints. Such models can help identify plans that will have higher benefit to cost ratios. Lai and colleagues have developed planning optimization models for the railroad industry. Once such model optimized the entire railroad network and selected links for capacity improvement for a given planning horizon (Lai & Barkan 2011). This model was subsequently adapted to analyze a single passenger rail line (Lai & Shih 2010) and select stations that would need additional platform tracks. More recently, it has been modified to select engineering projects that improve travel times between Nangang and Toucheng in Taiwan (Lai & Huang 2012a).

#### **2.2 Scheduling**

There are two main components of travel time for a train to travel from its origin to its destination: minimum-run-time (MRT) and slack. The MRT is governed by the physics of the train and railroad infrastructure. The train is limited by civil speed restrictions, and its own acceleration and braking performance. There are variations in railway operations that may cause

travel times to be greater than its MRT. It may encounter an opposing train on a single-track and stop on the passing siding to let the other train advance. It may have different power and tonnage causing different acceleration and braking performance. Another train may block the mainline for a mechanical failure. There may be temporary speed restrictions on the line due to maintenance activities.

A schedule should be based on a reasonable estimate of travel time including the MRT, and “slack” to account for the extra time due to variance in day-to-day operations. The MRT for a route can be calculated from a train performance calculator, or derived from field data. However, the amount of slack must be estimated. Figure 2-1 shows the travel-time broken down by MRT, slack, and unscheduled delays. In this example, the schedule accounted for the MRT and some of the variance in operations, but the slack did not account for all of the variance, such that there was unscheduled delay. Physics governs the MRT while capacity utilization and external factors will govern the amount of slack to include in the schedule. Often, the method of calculating on-time-performance metrics will cause schedulers to allocate slack at different locations of the schedule. A train schedule allows actual travel times to be measured against the operator’s expectation of travel time. In absence of a schedule, the actual travel time can be compared to MRT. Therefore there are two common methods of calculating train delays: the difference between total travel time and MRT, and the difference between total travel time and the schedule (Caughron 2013).



**Figure 2-1: Components of a schedule.**

There are two scheduling strategies that railroads can use: master scheduling and real-time scheduling (Hallowell & Harker 1998). Master scheduling is commonly used on European railroads. This involves developing a detailed timetable for scheduled trains and slots for unscheduled trains, and then operating with strict adherence to these schedules (Ireland et al. 2003). In real time scheduling, railroads use schedules to describe departure times from terminals, and provide guidelines to dispatchers for resolving conflicts between trains.

### **2.2.1 Passenger Train Schedules**

Many operations in Europe and Asia have strict on-time-performance requirements. In Switzerland 95% of the trains arrive at their destinations within 5-minutes or less of their scheduled time. These operations fall under “master scheduling”. Current practice in the United States is to schedule passenger trains very close to their minimum run time between intermediate stations. Most of the slack in the schedule is placed at the end-points of the route. This has the benefit of keeping pressure on the host railroad to preserve average train speed. Additionally, this

prevents the passenger train from arriving early and having to dwell on the mainline waiting for the scheduled departure time. As a consequence of strict schedules, the on-time performance of passenger trains is unachievable due to various deviations from the original schedule. Martland (2008) suggest “experienced based scheduling” where schedules are created to reflect the current operating performance of the inner-city trains. This will improve on-time arrival percentage and give the public more realistic expectations of travel time and punctuality. Martland acknowledge that actual travel times with the “experienced based schedule” should be similar to performance of the original schedule. Additionally, a policy should be in place to address what happens when a train arrives at a station earlier than the “experienced based schedule” (Martland 2008). This result of improving reliability by increasing the amount of slack in the schedule was repeated in simulations of the Stockholm to Goteborg corridor in Sweden (Sipilia 2012).

### **2.2.2 Freight Schedules**

With real-time scheduling, railroads use schedules to determine train departures times from terminals and crew changes. Dispatchers then make tactical decisions on how to route train movements over their territory instead of having the conflict resolutions planned in the timetable. This provides railroads with flexibility to adapt to fluctuations in freight traffic demands, as well as variability in train movements over long hauls and hand-offs between different carriers. Although North American railroads are becoming more scheduled, most traffic, other than passenger trains, does not conform to a precise schedule. For example, Canadian Pacific has adapted schedule for much of the manifest traffic (Ireland et al. 2003). These freight train schedules provide better service to the customers however the train performance has too much variability to institute “Master Scheduling”.

### **2.2.3 Temporal Separation of Traffic Types on Sharing Corridors**

In some cases, passenger and freight traffic can be separated by time of day with passenger trains running during the day and freight trains at night. This strategy has been implemented on some commuter rail corridors. For example, Westside Express Service in Portland, provides rush hour passenger service on tracks owned by the otherwise freight only Portland & Western Railroad (Leeson 2002). Additionally, Austin Metro and Denton County use local shortline railroad infrastructure to provide daytime service similar to conventional light rail systems while the owning railroad provides freight service at night. This solution is most effective when freight traffic is light and customer switching operations that do not require use of the mainline or need to be done during the hours of passenger operations. A challenge of executing this solution is that the existing infrastructure is often built for use over a 24-hour period. Depending on demand and schedule constraints, the freight service may be unable to be compressed to nighttime only operation. By limiting freight operations to night only, more than 50% of the line's capacity is taken away from the freight railroad. Additionally, the freight railroad may have customer service requirements that dictate some daylight operation (Bing et al. 2010).

### **2.3 Delays Due to Heterogeneity**

Railway traffic is considered homogeneous if all trains have similar characteristics. A good indicator of homogeneity is when most of the trains have the same average speed per track segment. Urban metro systems are the most homogenous rail systems in which all trains have similar characteristics such as running times and stopping patterns. Such systems achieve very high capacity (Pachl 2003) Conversely, in a mixed-use corridor, train types may include bulk

freight, high priority freight, commuter, and higher speed intercity passenger trains. Simple stringline diagrams show that heterogeneity in train types reduces line capacity (AREA 1921).

### **2.3.1 Single Track**

Bronzini and Clarke (1985) investigated North American operations using simulation to develop delay-volume curves for traffic with varying amounts of intermodal and unit trains on a hypothetical single-track line. They found that heterogeneity caused a substantial increase in train delay.

Jelaska (2000) developed a spreadsheet tool that created random schedules for a corridor featuring passenger and freight trains. Passenger trains were given preference and scheduled first. Then, freight trains were inserted in-between the passenger train paths. He calculated a linear relationship describing the tradeoff two train types. This was derived from the optimal frontier of the random timetables in which throughput of the line was maximized. Each point on this frontier included different combinations of passenger and freight trains. Jelaska was able to estimate this optimal frontier using linear regression and determined that each additional passenger train eliminated 1.25 freight trains.

Harrod (2009) modeled traffic using mathematical integer programming to consider the differing impact of mixing faster and slower non-conforming trains. He found that introducing a higher-priority, faster train will not reduce the total number of feasible train paths through the single track network. However, this will come at the cost of high delays to the slower, lower-priority train type. This work assumed that all sidings were long enough for meets to occur without stopping. Additionally, slower trains were always able to pull into a siding, thereby avoiding delays to the faster trains. Harrod suggests that assuming all trains must stop for meets (shorter sidings) will lower average speeds of the trains but not change throughput. Harrod



suggests that incorporating a stochastic schedules and delays are an area of future research (Harrod 2009).

Dingler et. al (2009) analyzed the interaction between intermodal and bulk trains. They simulated a single track line at various traffic volumes with different mixtures of intermodal and bulk trains. The delays followed the trend of an inverted parabola; average train delays were lowest when trains were most homogenous and greatest in heterogeneous mixtures. The authors suggested that the delay due to heterogeneity was caused by the intermodal trains travelling at greater speeds, higher dispatching priority, and a higher horsepower-to-trailing-ton-ratio permitting more rapid acceleration. They systematically investigated the delay attributed to heterogeneity in train type by isolating each of these differentiating factors. Equalizing the priorities between trains had the greatest effect on reducing train delays, followed by acceleration performance. The paper also described the interaction between passenger and freight trains. Train volume was held constant and passenger trains replaced freight trains on a 1-to-1 ratio. Each additional passenger train round trip increased freight train delays but with little delay to the passenger trains. The authors suggest a linear relationship between freight train delay and the number of replacement passenger trains.

Gorman (2009) created a train-run-time model using empirical data from eight BNSF Railway subdivisions. He identified meets, passes, and overtakes as the principal causes of delay. Dingler et al. (2010) expanded on Gorman's work using simulation analysis to analyze delays between intermodal and bulk trains and categorizing the delays by conflict type and source (Table 2-1). Identifying the operational sources allowed the authors to determine which delay conflict produces more delay and why that delay is occurring. Meet delays were caused when one train was delayed due to a conflict with one or more trains in the opposing directions.

Reduced speed related delays were minor and relatively constant. These types of delays occurred when the signal system forced a faster train to pace behind a slower train (approach aspect). Passes were not a major source of delay on single track, indicating that speed difference alone was not a major factor. Trains stopped on sidings waiting for a meet were the leading cause of delay (Figure 2-2). Delays due to braking and accelerating increased as the percentage of bulk trains in the traffic mix increased, making these delays more dependent on the type of train than on the heterogeneity of the traffic. Delays when a train was stopped were the only source of delay that increased with heterogeneity (Figure 2-3). Therefore, heterogeneity increased delay by increasing the time that trains were stopped on sidings. Dingler et al. provided two possible explanations for this result. First, at the higher levels of heterogeneity there was a greater likelihood that two trains of different priorities would meet, resulting in less efficient meets with more time stopped. Second, higher levels of heterogeneity resulted in more complex conflicts in which a train was met or passed by more than one train, resulting in more time stopped on the siding.

**Table 2-1: Categories of Delays.**

Conflicts	Sources
Meets	Accelerating
Passes	Braking
Line	Reduced Speed
	Stopped

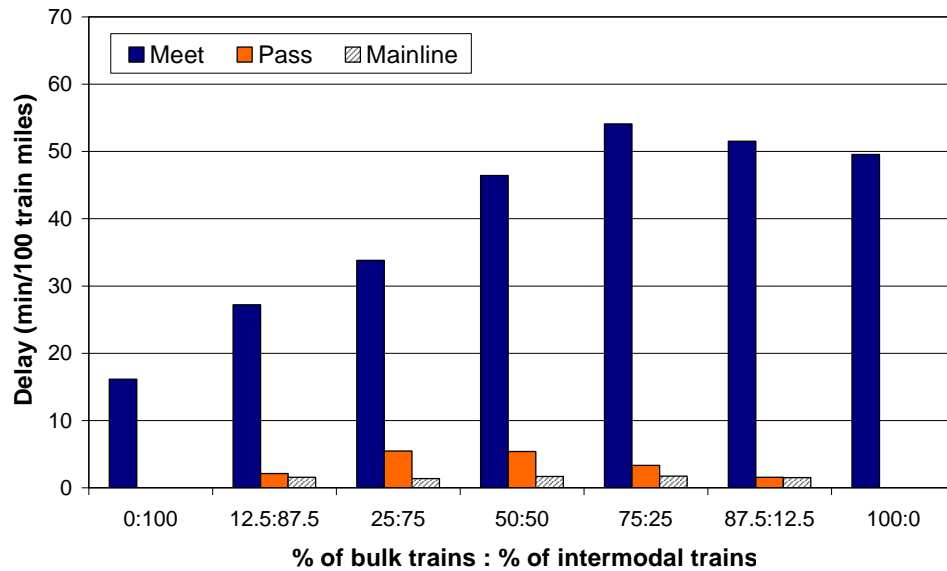


Figure 2-2: Train delays by traffic composition grouped by conflict type.

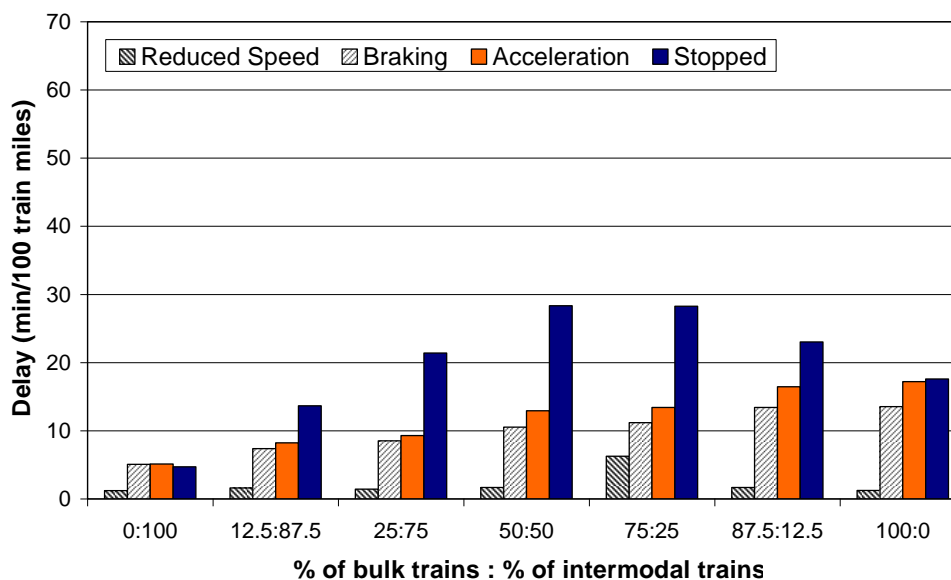


Figure 2-3: Train delays by traffic composition grouped by "source".

### 2.3.2 Multiple-Main-Tracks

Abbott (1975) analyzed the interaction of 50-mph freight trains operating alongside passenger trains traveling at 80-mph, 125-mph, and 150-mph. Abbot only considered one direction of a double-track line, where the passenger trains have absolute priority over freight trains. Abbott reported that through freight paths decline from the base (no passenger) scenario by 28%, 57%,

and 71%, respectively, and that freight train run-times increased by 33%, 58%, and 88%. However, because the trains all travel in a single direction, no meets between opposing trains occurred, only passes.

Vromans et. al. (2006) used simulation to study heterogeneous passenger services and developed measures of heterogeneity. By giving local and long distance trains the same number of station stops the authors were able to homogenize the train schedule and improve operations. Abril et. al. (2008) used simulation to investigate different factors influencing capacity on Spanish rail lines. One of the factors considered was trains operating at two speeds on double-track lines: “normal” and 50% of normal. Results showed capacity declined as heterogeneity of train speeds increased (Abril et al. 2008).

Lindfeldt (2012a) simulated double-track mainlines with passing sidings that can be used for faster trains overtaking slower trains. His simulations included express intercity trains, local commuter, and freight trains. Lindfeldt conducted a sensitivity analysis of the distance between overtake points and the traffic composition. He found that the distance between overtake locations had a minor effect compared to heterogeneity in speed between the services in the timetable. Having heterogeneous operations consumed most of the slack in the schedule (Lindfeldt 2012a).

## **2.4 Influence of Infrastructure**

In order to add traffic to an existing corridor, the host railroad would likely require any additional delays to be mitigated and pay for the capacity consumption. Two common techniques in delay mitigation are adding additional tracks to the railroad line or changing the operation to be more efficient (AREA 1931).

Petersen and Taylor state that an important way of mitigating delay due to meets is lengthening passing sidings. They describe an optimization method for locating passing sidings such that opposing trains are able to pass each other without coming to a full stop, typically referred to as a “rolling meet.” From these train characteristics and information about headways and travel times, they determine the number and location of meets and the length required for each siding. This scenario is solved for the various speeds allowed on the siding and the various cruising speeds. Next, the authors determine the ability of the designed track to resist delay or, when delay occurs, avoid compounding that delay. They refer to this as the “robustness” of the track design. They found that increased length of sidings, as well as “schedule padding,” both contribute to the robustness of track designs. They show that “extra siding length has a strong impact on the ability of the line to recover from unexpected delay. The authors suggest that 13% to 20% partial double tracking may achieve performance similar to full-double-track assuming strict adherence to the meet-pass planning (Petersen & Taylor 1987).

There has been limited research on hybrid-track configurations. Pawar (2011) used analytical models to determine the length of long sidings needed to run a single-track, high speed railway without delays due to meets. They suggested a linear relationship between trains per hour and the percentage of double track. Additionally, the flexibility of the timetable was considered by extending sidings an additional amount, increasing the likelihood of a “rolling-meet” occurring. The amount of track needed was derived from the slack time incorporated into the timetable. Greater slack times increased the double-track percentage to ensure a rolling-meet occurred on the long sidings. Lindfeldt (2012b) compared partial double track to additional sidings with simulation techniques. Many possible timetables of different traffic levels and sequencing of trains were considered in the analysis. The capacities of the partial double-track

and additional siding scenarios were similar; however, there were more feasible timetables in the partial double-track scenario. This would indicate that partial double track may be more robust to resisting delays.

## CHAPTER 3

### RAILROAD CAPACITY METRICS

#### 3.1 Introduction

The North American railroad network is expected to experience continued growth in freight traffic. Overall freight transportation demand is projected to increase 84% by 2035 (AASHTO 2007). Freight railroads continue to invest in intermodal freight cars and terminals. The strongest growth is in intermodal freight and traffic related to petroleum, liquefied natural gas and ethanol products. Meanwhile new regional and intercity passenger services are being proposed to operate over portions of the freight railroad infrastructure. The combined growth of these two different types of rail traffic is expected to tax the capacity of many rail lines. However, quantitative measures of the impact of this new traffic can vary considerably based on the capacity metrics used. Additionally, interpretation of these metrics, and even the term “capacity” itself, can have different meanings to different stakeholders. By contrast, highway transportation planners, practitioners, and researchers have benefited tremendously from the development of the Highway Capacity Manual and its consistent set of metrics, applications and interpretations. No comparable manual is available to provide an objective scientific foundation for rail capacity planning. This chapter will derive key railway traffic relationships and measurements that could potentially be used to help standardize measures of railway capacity and improve communication and comparison of railway capacity metrics among various stakeholders. Additionally, this chapter will compare established highway traffic relationships and develop the railway corollaries.

The core of all transportation statistics is describing the trajectory of a vehicle. Physically, these are described by time, distance, and velocity. A significant challenge of

transportation operations management is not describing one trajectory but instead describing the quality of the entire service. There are numerous metrics used by both the rail and highway modes to analyze and plan operations. A subset of these will be examined in this chapter. Additionally, railroads and highways have certain key traffic relationships that will be compared. The concepts presented in the first half of this chapter will then be illustrated by simulating a hypothetical railroad line. Understanding measures of vehicle performance and reliability as metrics for capacity can help guide decisions on service improvement.

### **3.2 Differences between Highways & Railways**

Table 3-1 illustrates some of the fundamental differences between rail and highway modes of transportation. Generally, highways are more analogous to a fluid than a railroad, and this is reflected in their respective capacity configuration. A low-capacity, baseline highway may be a two-lane rural road where traffic is separated by direction. Overtakes between faster and slower vehicles are possible provided there is sufficient sight distance and no on-coming vehicles. On a railroad, a low-capacity baseline would feature single track with passing sidings. Trains can travel in opposite directions on the same track at different times. When the trajectories of two trains intersect, the theoretical conflict point, then the train paths need to be modified such that this interaction occurs at a passing siding. The spacing of the passing sidings governs where conflicts can be resolved between trains. In high capacity configurations, a highway may feature two or more lanes in each direction with express or HOV lanes. Faster traffic can overtake slower traffic without conflicting traffic traveling in the opposite direction. A high capacity railroad may feature two-main tracks separating direction of travel. Overtakes can only occur at discrete locations of crossovers where trains can change tracks. Additional sidings or main tracks



may be necessary to perform overtakes without delaying traffic in the opposite direction. Even a low-capacity, two-lane highway theoretically has an “infinite” amount of crossovers where vehicles can overtake. Other major fundamental differences between highway and rail are that trains have a much longer stopping distance, slower acceleration, and that most traffic can be strategically managed by a central dispatcher. Despite these differences, highways have traffic relationships, and performance metrics that can complement railway capacity analyses.

**Table 3-1: General characteristics of rail and highway modes of transportation.**

	<b>Railroad</b>	<b>Highway</b>
Number of vehicles per day	1~300	1~200,000
High capacity configuration	2 main tracks with crossovers spaced at 5-10 mile intervals with additional main or siding tracks to execute overtakes at discrete locations	At least 2 lanes in each direction with possible express or HOV lanes.
Stopping distance	1~3 miles	15~300 feet
Acceleration	0 – 60 mph measured in minutes	0-60 mph measured in seconds
Speed control	Signals at discrete locations	Relative to downstream vehicle
Traffic management	Central dispatcher	Ramp metering or congestion toll
Terminals	Classification yards, industrial sites, and stations	Driveways and parking lots
Infrastructure owner	Usually private (North America)	Usually public
Vehicle owner	Freight – Private & Passenger- Public	Private

### 3.3 Types of Capacity

The capacity of transportation infrastructure is often measured as the maximum flow rate of some quantity (volume or mass) of substance passing a point per unit time. Kahn (1979) offered four important definitions of railroad capacity in a study conducted for Transport Canada:

1. *Practical Capacity*: Ability to move traffic at an “acceptable” level of service
2. *Economic Capacity*: The level of traffic at which the costs of additional traffic outweighs the benefits

3. *Engineering Capacity*: The maximum amount that can possibly be moved over a network (Theoretical Capacity)
4. *Jam Capacity*: The system has ceased to function and all trains are stopped

These four capacity definitions are listed in order of magnitude from smallest to largest. The first and second definitions of capacity should be approximately equal. Rail lines should continue to serve new traffic until the costs of congestion are greater than the benefit of additional traffic. *Practical Capacity* is achieved at the traffic level where the service quality has deteriorated to a minimum level of service (MLOS) that is still acceptable where additional traffic violates this standard. This standard is set by the infrastructure based on the type of traffic using the railway line. For example, a passenger operation will usually have a stricter on-time performance requirement than a freight operation, and be more sensitive to congestion related delays. *Economic Capacity* is determined by calculating the traffic level where the actual costs of congestion equals the revenues of additional traffic. For a railroad, costs of congestion and lost revenue are borne by a limited number of parties so operating beyond this capacity is rare. Conversely, highways are not governed by the same revenue relationships as railroads, because operating costs are distributed over many users, making *Economic Capacity* harder to define on the non-tolled, highway network. Demand can exceed *Economic Capacity* causing throughput to deteriorate. Operators of tolled highways can use traffic management techniques to mitigate the externality of congestion by dynamically adjusting the tolling rate in response to traffic levels. Additionally, ramp metering can be implemented to maintain traffic flow.

*Engineering Capacity* is the maximum flow through the network and often corresponds with high variability. It can be described using analytical techniques for both railroads and highways. The *Jam Capacity* can also be determined analytically, but this value is no longer a flow measurement because there is no movement. This condition is often described as a linear

vehicle density. Highways often exceed the *Engineering Capacity* and can temporarily reach a *Jam Capacity* condition. On a railroad, traffic is generally managed so as to not exceed the *Practical Capacity* of a line (Kahn 1979). A well-documented exception was the situation that developed after the Union Pacific merged with the Southern Pacific in 1996 (Kwoka & White 1997).

### 3.4 Traffic Flow

Highway capacity measurements are based on three key elements: speed density and flow. A general metric for the first is average velocity of vehicles through the highway network. Second, is the density of the vehicles on the highway measured by the number of vehicles per mile. Lastly, the flow of the highway is measured by the number cars passing a given point per unit of time. Velocity and flow are network health metrics used by both highways and railroads. Velocity is measured as the average vehicle speed for both modes. The highway mode refers to flow as either hourly vehicle throughput or average daily traffic. These three metrics are related to each other by Equation 3-1. This equation is analogous to the volumetric flow equation in fluid mechanics where cross-sectional area is replaced with the linear density of vehicles in the traffic “pipeline” (Transportation Research Board 2010).

$$q = kv \tag{3-1}$$

Where:

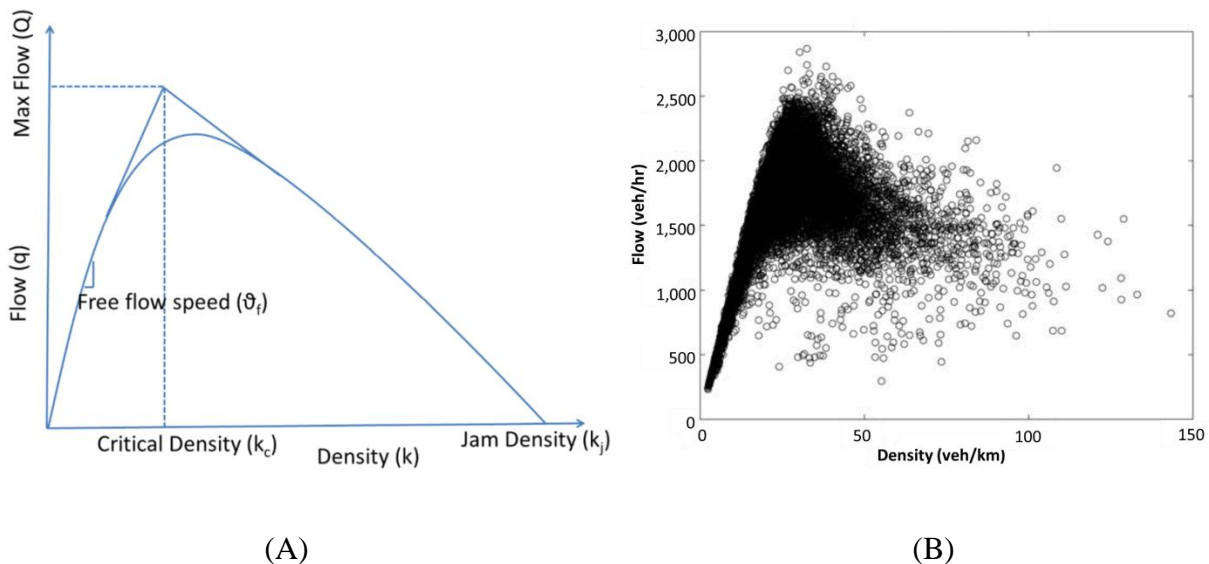
$q$  = Flow rate (vehicles/hour)

$k$  = Linear density (vehicles/mile)

$v$  = Average velocity (miles/hour)

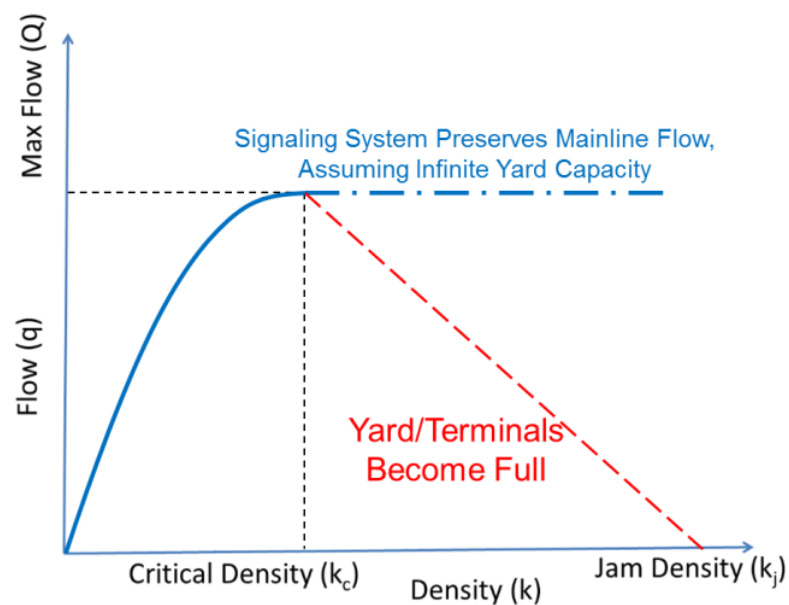
Vehicle interactions and safe following distance between vehicles can constrain flow in dense highway corridors. This degradation in traffic flow is characterized by the fundamental highway traffic diagram (Figure 3-1a). This diagram indicates that as more traffic enters the

highway, the vehicle flow through the highway increases until a critical density is reached. The maximum flow of vehicles through the highway occurs at this point (Greenshield 1935). Maximum flow is considered to be the capacity of the highway and would correspond to *Engineering Capacity*. After this point, additional vehicles entering the highway will reduce the flow through the system. The downward sloping section of the diagram corresponds to increasingly congested travel characterized by slower speeds and greater variation in speed (stop and go travel). Ultimately, the jam condition is reached where all vehicles on the highway are stopped and there is zero flow of vehicles through the network. This condition corresponds to the *Jam Capacity* (Kahn 1979). The transition from increasing traffic flow to decreasing traffic flow is actually a more stochastic process than suggested by the Greenshield model as seen in Figure 3-1b. An important consideration is that for any traffic flow rate, there are two corresponding vehicular densities. The lower density corresponds to a free flow speed environment and the high density corresponds to a congested environment.



**Figure 3-1: (A) Fundamental traffic diagram and (B) actual flow-density data on a highway corridor.**

Railroad transportation differs from this highway traffic diagram due to some unique characteristics of railroad operations. Most importantly, railway traffic is much more discrete and less analogous to a fluid in a pipeline or a continuous stream of highway vehicles. Secondly, railway traffic is managed by a dispatcher and safe train movements are governed by a signal system. Under wayside automatic-block signaling, the locomotive engineer is given information every two to four miles about the speed and track occupancy necessary to prevent a collision with a train ahead of it. The three typical signal aspects are clear, approach, and stop whose indications mean proceed at normal speed, approach the next signal prepared to stop, and stop, respectively. A potential corollary fundamental traffic diagram for a railroad is shown in Figure 3-2.



**Figure 3-2: Applying the fundamental highway diagram to a railroad subdivision.**

On a double-track line, the free flow speed is maintained until the train density results in train spacing equal to the signal spacing multiplied by the number of blocks ahead required to receive a clear proceed signal. After this critical density, the flow rate is maintained by the signal

system, and trains must wait for a clear signal and queue in a terminal similar to highway ramp metering. If there is near infinite terminal capacity to store trains, then rail traffic density can increase without flow deteriorating. However, this infinite terminal capacity assumption is unrealistic. Eventually, capacity constraints within the terminal will prevent trains from entering and leaving the corridor and degrade traffic flow. For a highway, the degradation in flow is due to vehicle interactions and for the railroad the degradation in flow is due to terminal constraints.

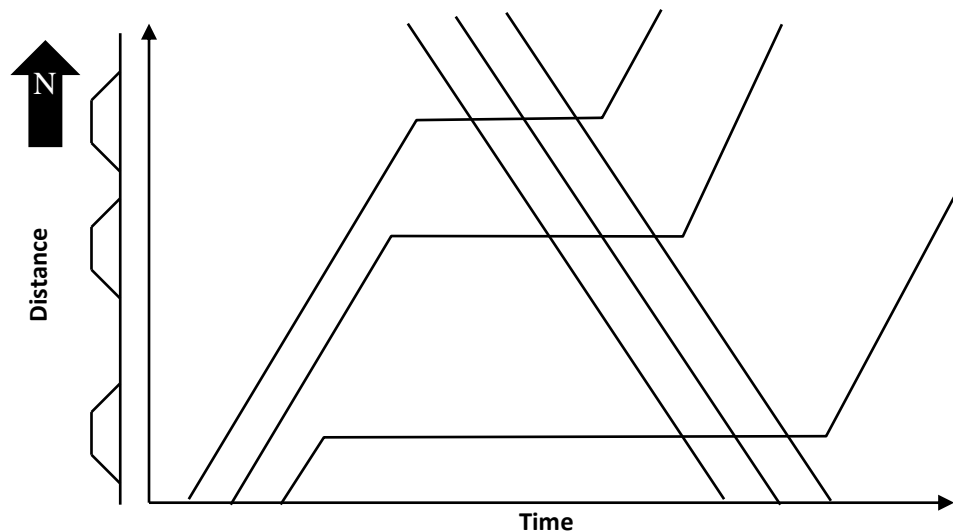
The depiction of Figure 3-2 for a single-track line is justified by slightly different reasoning than for double track. In the case of trains traveling in opposite directions on a single-track corridor, there must be sidings to allow trains to pass (Figure 3-3). In congested conditions, trains stop at passing sidings and take alternate turns using the single-track sections between passing sidings. In this case, the flow of traffic is not restricted by the signaling system, but by the process of trains taking turns using each bottleneck section.



**Figure 3-3: Example single track railroad line.**

The maximum number of conflicting trains at one location is generally three trains. For example, a dispatcher can hold an eastbound train at a siding and let two westbound trains go by it before continuing. If a group of eastbound trains are in conflict with a group of westbound trains, then this conflict is usually split between separate sidings. However, use of adjacent sidings may already be needed to resolve other conflicts between different sets of trains. This can cause trains to wait for extended periods, leading to a traffic flow breakdown. Figure 3-4 is a time-distance diagram illustrating two groups of trains (fleets) traveling in opposite directions.

The northbound group is separated by trains pulling into sidings to let the southbound trains go by. The last northbound train occupies the siding for the longest period of time. The northbound fleet is broken up and occupies the sidings for extended periods of time thereby reducing the traffic fluidity of the corridor (Kraft 1982). Depending on the siding spacing and the directionality of the traffic, there exists a critical density of mainline traffic where the flow will start to breakdown in this manner. This is caused by saturation of the mainline passing sidings. However, the traffic is centrally managed, such that flow can be maintained by holding trains in terminals similar to the double track analysis. For single track, degradation in traffic flow can result from either filling terminal capacity or causing too many complex interactions at the bottleneck sections.



**Figure 3-4: Adopted from Kraft (1982); Northbound Trains pull over in sidings to let the southbound trains continue un-impeded. This separates the northbound fleet.**

### 3.5 Delay Relationships

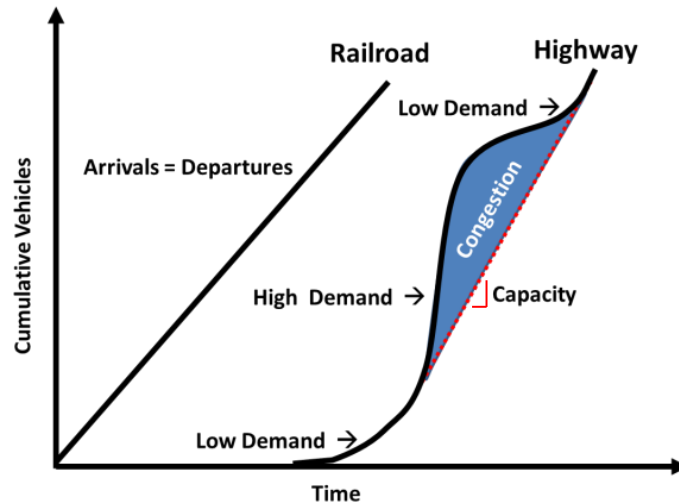
There are two broad categories of improvements that can be made to a transportation network. The first are process improvements that reduce the amount of time that a particular resource or

asset is used. For example, increasing the speed limit reduces the amount of time a vehicle is using its guide-way and can free up that resource for the next vehicle. The other type of improvement is queuing capacity. Queues are where vehicles wait to use the next infrastructure resource in the transportation process. A left turn lane is a queue where automobiles wait to use the intersection resource. If the left turn lane is too short, then the queue to turn left can extend into the traveling lanes and interfere with the traffic traveling straight. Smart traffic lights that sense vehicles can do a better job of managing the flow of vehicles through an intersection and improve process times. Adding more receiving tracks to a classification yard is a queue improvement such that more trains can wait to be classified. Adding automatic doors to passenger trains reduces dwell times and improves processing time at stations. Adding a siding to a railroad is both a process and queue improvement. Acting as a queue, the siding allows for a train to wait and use a single track resource. Additionally, acting as a process improvement, a siding allows for interaction between trains to occur at an additional location closer to the theoretical conflict point of the train trajectories. This results in improved utilization of the single track bottleneck resource.

An important consideration is matching the desired capacity to the expected demand of the transportation infrastructure. Building at a greater capacity may cost more to build and maintain but can also be more capable of handling demand surges. Railroad and highways build different relative amounts of infrastructure. Railroads can build and maintain less route infrastructure because it is more closely matched to a near uniform traffic flow demand to improve asset utilization and minimize infrastructure construction and maintenance costs. Highways are designed to accommodate the high demand periods during the morning and evening rush hours. When a period of congestion does occur on a highway, the queue of traffic



can overlap into a period of low demand and the highway can transition back into the free-flow condition (Figure 3-). However, under congested railroad conditions, there is often not a time period of low demand to relieve congestion within the network. However, in a mixed-use corridor with passenger and freight train, there may be demand spikes. Freight railroads that share track with commuter operations often cannot operate freight trains during the peak rush hour periods. The type of analysis in Figure 3-5 can be used by freight railroads to determine the length of a freight train curfew and the size of any holding yards that may be needed to accommodate passenger trains in rush hour.

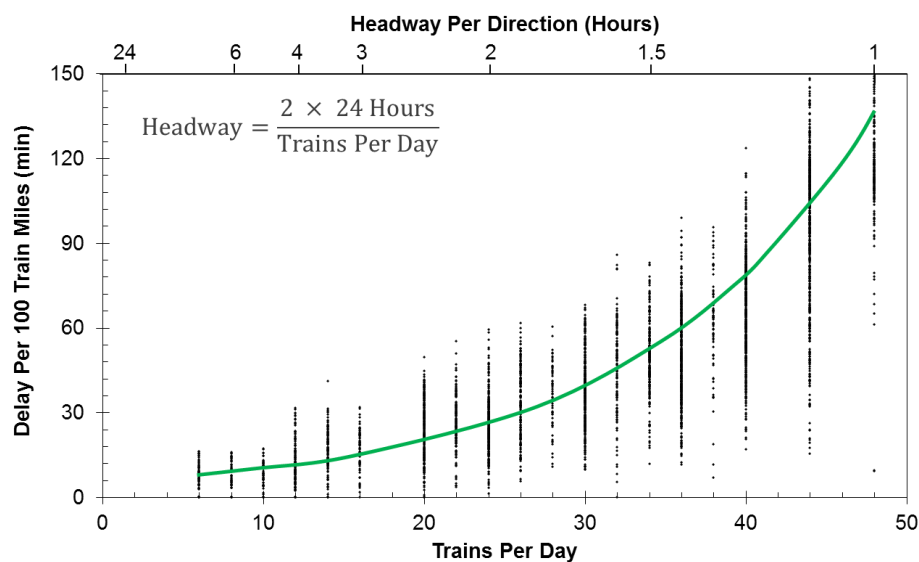


**Figure 3-5: Cumulative arrival and departure curves for (left) railroad and (right) highway. When demand exceeds capacity, vehicles must “queue” to use the infrastructure. Railroads can more closely match capacity to demand they can centrally manage their traffic.**

The delay of each train is calculated by the difference between the actual run time and the minimum-run-time (MRT) of the train. The MRT is the minimum time that a train can traverse a route without any interactions with other trains. This calculation of train delay is independent of any scheduling assumptions. A schedule or timetable for the trains can be derived from distributions of these raw-delay values by incorporating buffer time (slack). Certain delays can be expected to occur such as the meet delay at a siding between two passenger trains and these

can be incorporated in to the schedule. These “scheduled delays” can be included in the buffer time to create a more predictable operation for railroad users.

A key relationship with railroad traffic is that higher traffic densities lead to higher delays. Each data point in Figure 3-6 represents the delay of one freight train. Without any improvements to the infrastructure, greater traffic levels correspond to higher delays with a broader range of run times. This curve shifts to the right if more track is added or if network efficiency increases. Unlike the fundamental traffic diagram, a delay volume curve does not yield a discrete flow measurement for the capacity of the line. However, further analysis can allow the planning department of a railroad to devise a capacity value from this curve. A simple method of defining the capacity is the maximum number of trains per day where a minimum level of service (MLOS) is still maintained. The MLOS definition will differ for different operators and traffic types. One method might be to state a maximum allowable average delay incurred by trains on the network, and extrapolate the corresponding volume from the curve in Figure 3-6



**Figure 3-6: Delay volume relationship of railroad traffic.**

The delay-volume curve can be used to define the capacity, given a level of service constraint or a change in capacity between two similar lines. This can be accomplished by defining the mathematical relationship describing the delay-volume curve through regression techniques. Usually an exponential relationship is used (Equation 3-2a) although polynomial models also fit well (Krueger 1999, Abril et al. 2008). If a railroad defines MLOS by a maximum tolerable average delay,  $D_{max}$ . Equation 3-2a can then be rearranged and solved for railroad capacity (Equation 3-2b). Consider two different single-track routes that have different amounts of double-track sections installed. The delays experienced on each of these two routes could be explained by Equation 3-2b with a unique value of  $A$  for each route. If each route is operating at the MLOS,  $D_{max}$ , then Equation 3-3 can be used to calculate the difference in capacity of the two lines. The difference in capacity of these two lines is independent of the exact MLOS value. Equation 3-3 is only valid if the traffic and operating characteristics between the two lines are similar.

$$D = Ae^{kv} \quad (3-2a)$$

$$V = \frac{1}{k} \ln\left(\frac{D_{max}}{A}\right) \quad (3-2b)$$

Where:

$D$  = Train delay

$V$  = Traffic volume

$A$  = Route constant

$k$  = Delay growth constant

$$V_2 - V_1 = \frac{1}{k} \ln\left(\frac{A_1}{A_2}\right) \quad (3-3)$$

A disadvantage of delay as a capacity metric is that it is no longer a fair comparison when the MRT is changing between alternatives. In this case, the distance lost per minute of delay is no longer constant. Additionally, there can be improvements to railroad infrastructure that reduce

run-time but leave delay unchanged (White 2006). However, using travel time as a metric for capacity improvement projects is confounding because it cannot distinguish between reductions in travel time due to a lower MRT and reductions in travel time due to less congestion. Additionally, regression analysis of travel time and traffic volume will show very high fit statistics for a linear model. An analyst may easily overlook evidence in the residuals indicating that a curvilinear model is necessary.

Average train speeds can vary on a railroad line. A heavy bulk train may operate at 30-mph on a line that also has passenger trains traveling at 79-mph. Heterogeneity in train type increases the number of train interactions and consequent delay (AREA 1921, Dingler et al. 2009). Similarly, personal vehicles and heavy trucks consume different amounts of highway capacity due to their differing operating characteristics. Delay-based efforts have been able to equate different vehicle types into Passenger Equivalent Units (Benekohal & Zhao 2000). There has been research attempting to equate capacity consumption to train type. Lai et al. (2012) proposed a Base Train Equivalent Unit (BTE) that is defined as the delay ratio of a marginal non-base train over a base train. Different train types can be assigned a BTE based on route characteristics in order to standardize the capacity consumption measurement of that particular line. For example, a 40-mph freight train may be the standard unit with a BTE of 1.0 and a 79 mph passenger train may have a BTE of 3.0. A line with 30 standard freight trains requires 30 train units of capacity. A shared corridor with 20 freight and 10 passenger trains would consume 50 train units of capacity. Calculation of BTE can vary depending on the operating characteristics of the trains and route. Lai et al. (2012) determined the BTE between coal and intermodal trains under various operational conditions in order to setup a BTE database.

Calculating BTE values from a database is not a substitute for simulation analysis but rather an aid to help planners understand how heterogeneity in train types increases capacity consumption.

### 3.6 Transportation Economics

Transport Canada's second definition of Economic Capacity is illustrated in Figure 3-7. The capacity of the line is determined to be at the point where the marginal costs of additional traffic equal the marginal benefits of that traffic. The marginal cost curve is derived from costs both independent of, and dependent on congestion. There are fixed cost for running an additional train such as fuel, labor, locomotives, and railcars that vary minimally with increased traffic levels. At low traffic densities, these costs dominate the calculation of marginal cost. At higher traffic levels, there is more delay, and the cost of congestion dominates the marginal cost calculation (Figure 3-7). In the calculation of delay costs, there are immediate direct costs to a railroad such as increased fuel consumption, shipment-contract penalties and potential crew expiration costs.

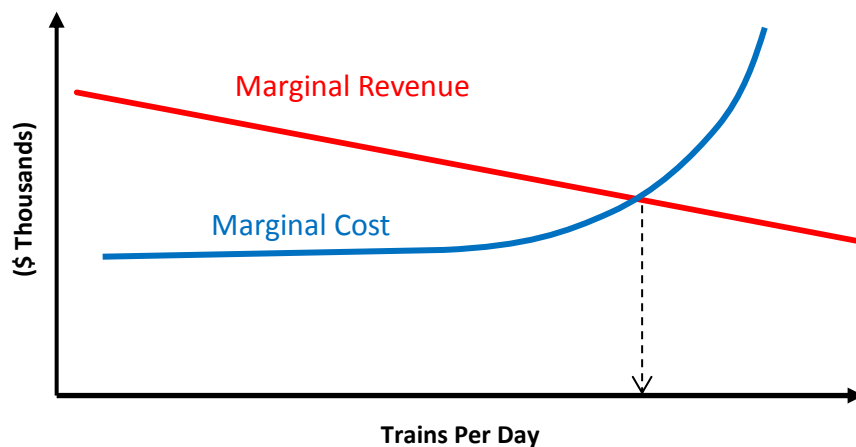


Figure 3-7: Economic Railroad Capacity.

The indirect costs of high delays are poor asset utilization. Slower velocities can lead to more crews, locomotives, and railcars needed to satisfy a longer cycle-time (Oh et al. 1995). This

loss in asset utilization can be characterized by multiplying the average speed of the MRT by the delay of the trains. This is an opportunity cost that can be valued as the extra distance the railroad could have moved shipments in the same time. For example, if a train's average speed corresponding to its MRT is 60-mph and this train is delayed 60-minutes, then the opportunity cost of that delay is an additional 60-miles (Equation 3-5). The loss of 60-miles of potential goods transportation during the delay time translates into lost transportation production and reduced transportation efficiency. If the route is 200-miles long, then the transportation efficiency of the route is 77% (Equation 3-6a). The transportation efficiency is a measure of the fraction of theoretical free-flow transportation production that can be achieved given train interactions and delays. This value is the complement to a more common metric, delay percentage (Equation 3-6b). This efficiency loss can exist in an economic equilibrium where the opportunity cost of this delay is less than the cost of installing the amount of track that would alleviate delay from the system.

$$X_L = \forall_{MRT} D \quad (3-5)$$

Where:

$$\begin{aligned} X_L &= \text{Distance not traveled due to delay} \\ \forall_{MRT} &= \text{Average speed of minimum run time} \\ D &= \text{Average train delay} \end{aligned}$$

$$e = \frac{X_R}{X_R + X_L} = \frac{\forall_R(MRT + D)}{\forall_R(MRT + D) + \forall_{MRT} D} = \frac{MRT}{MRT + D} \quad (3-6a)$$

$$\delta_{\%} = \frac{D}{MRT + D} = 1 - e \quad (3-6b)$$

Where:

$$\begin{aligned} e &= \text{Transportation efficiency} \\ X_R &= \text{Route distance} \\ \forall_R &= \text{Average speed of actual run} \\ \delta_{\%} &= \text{Delay percentage} \end{aligned}$$

The marginal revenue curve can be assumed to be horizontal when the price of shipping cargo in a railcar is assumed to be equal to the railroad's marginal revenue per railcar. This assumes that the railroad industry is perfectly competitive, where the price of shipping a unit is independent of an individual railroad's decision-making process. The other extreme is the monopoly condition where the railroad has the market power to set their own prices. Under this condition, the marginal revenue curve is downward sloping with a slope that is twice as steep as the demand curve. This indicates that the railroad has the ability to lower prices in order to attract new traffic (Varian 2006). In practice, the market condition of the railroad will often be somewhere between monopoly and perfectly competitive depending on geographic conditions or the commodities in question. Overall, this type of economic analysis requires data from a variety of sources. The fixed costs and revenues per train type are simple calculations. However, the cost per delay hour and the degree of market power to set prices are much more complex calculations. Calculating economic capacity may not always be feasible if data are not available.

Another approach to determine capacity is to implement a utility model to estimate the capacity of the line. This utility model incorporates the tradeoff between running more trains and reducing the average velocity. The capacity utility model will have a clear maximum value that indicates the capacity of the line (Pachl 2009). The maximum of the equation is dependent on the values of the exponents  $a$  and  $b$  that a planner chooses. The ratio between these two exponents indicates the marginal rate of substitution between running more trains and decreasing average velocity. Using 1 for  $a$  and 2 for  $b$  would make the utility model analogous to the formula for kinetic energy  $\left(\frac{1}{2}mv^2\right)$  where the number of trains is assumed to be the mass,  $m$ , and  $v$  is assumed to be average velocity of those trains. If these values are used for analysis of the line,

then the capacity of the railroad line could be measured in physical units of energy such as joules.

$$U = N^a V^b \quad (3-4)$$

Where:

$U$  = Capacity utility

$N$  = Number of trains per day

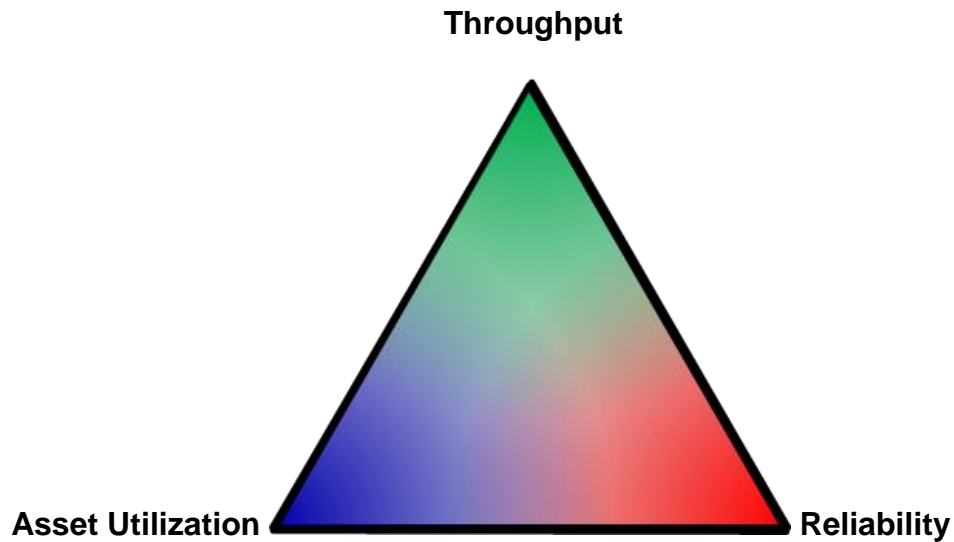
$V$  = Average train velocity

$a, b$  = Exponents chosen by planner

### **3.7 Performance Metrics**

Practical and Economic Capacity are derived from the quality of service provided. There are three interrelated broad categories to measure quality of service or capacity of a transportation system (Figure 3-8). The first is the throughput, the amount of goods or people that the transportation network serves per unit of time. The second is reliability of the service provided by the transportation network. Lastly there is the overall utilization of the existing assets. These three categories are interdependent. For example, having higher reliability may require operating with a lower throughput (AREA 1931). Having high asset utilization may result in periods of time with unavailable resources effecting reliability and throughput. An infrastructure improvement project can be undertaken to increase throughput as well as increase reliability but by definition decreases asset utilization. Often, capacity projects can be postponed if more efficiency can be gained using the current physical infrastructure thereby improving asset utilization (Sogin et al. 2012a).





**Figure 3-8: Level of Service Interrelationships.**

Example railroad capacity metrics are summarized and grouped by the categories in Table 3-2. The amount moved depends on the type of operation. Freight railroads might be more interested in measuring revenue-tons per day, while a commuter railroad might be interested in the number of passengers transported per hour in the morning peak. The reliability metrics describe the distribution of run time. Asset utilization can be measured by the amount of time that assets are (or are not) generating transportation activity and revenue.

**Table 3-2: Classification of Railroad Performance Metrics.**

Amount Moved	Reliability	Asset Utilization
Trains	Distribution of Arrival Times	Dwell Time in Terminals
Cars	Average Train Delay	Blocking Time
Revenue Tons	Standard Deviation of Train Delay	Signal Wake
People	On Time Performance	Train Miles/Track Mile
TEUs	Crew Expirations	Idle locomotives
(per unit of time)	Average Velocity	Cycle Time

Common highway metrics are identified in Table 3-3 (Biehler et al. 2008). The delay per traveler is analogous to the delay per train. The Travel Time Index is similar to a measurement of average delay. The Travel Time Index expresses the delay as a percentage of free flow time such that different roads can be compared regardless of their speed limits. Using the travel time index in delay analysis can mitigate the issue of delays having different consequences for trains with different speeds. The Planning Time Index and Buffer Time Index are metrics that measure the variation of travel times. The Planning Time Index measures the range of the travel time distribution by comparing the 95<sup>th</sup> percentile of travel times to the free flow travel time. The Buffer Time Index determines the amount of additional time that should be added to the average trip time in order to have an on-time arrival rate of 95%. Consider a case where the free flow trip time is 20 minutes, the average trip time is 30 minutes, and the 95<sup>th</sup> percentile of trip times is 50 minutes. In order to be on time 19 out of 20 trips, the vehicle must add 20 minutes of slack time to its schedule and depart 20 minutes early. The Planning Time Index is 2.5 and the Buffer Time Index is 67. The Planning Time Index indicates that a likely slow travel time is 2.5 times greater than the free-flow speed. The Buffer Time Index indicates that 67% of the average trip time should be added as schedule slack to maintain a 95% on-time arrival percentage (Sogin et al. 2012a).

**Table 3-3: Common Highway Metrics.**

Metric	Description
Velocity	The speed vehicles traveling through a network
Flow	The amount of vehicles passing a given point per unit of time
Density	The number of vehicles per unit length of infrastructure (railroad or highway)
Delay Per Traveler	The amount of additional time per person to reach a destination when compared to the free flow speed
Travel Time Index	The ratio of average trip time to free flow trip time
Planning Time Index	The ratio of the worst case likely scenario to the best case scenario
Buffer Time Index	The amount of slack time necessary in a passenger's schedule to have 95% on time performance
On Time Arrival Percentage	The percentage of vehicles arriving at their destination within a standard length of time
Volume to Capacity Ratio	The ratio of the current traffic levels to the maximum capacity of the highway
Misery Index	The ratio of the average of the slowest travel times in the 20 <sup>th</sup> percentile to the average travel time.
Level of Service	A qualitative grade (A-F) describing the capacity conditions of the highway as defined by the AASHTO Highway Capacity Manual

Measuring on-time arrival percentage (OTP) is another method of analyzing the reliability of highway vehicles as well as trains. This reflects the percentage of arrival times that are within a specified time interval. This is a common metric because it is easy to calculate and to understand by outside parties. However, the standard of what is “on-time” and what is “late” is set by the operating agency and this standard can be manipulated. In 2008 Amtrak increased the Auto Train’s scheduled trip time by one-hour and improved OTP from 10% in 2006 to 82% (Martland 2008). Thus, OTP cannot be used as a comparison tool between traffic types. For example, a passenger operation may have an OTP of 95% where on-time is considered to be arriving within 15 minutes of the schedule. A freight operation may also have an OTP of 95%, but “on-time” is within 3 hours of the train’s schedule. For a manifest freight train, not all

railcars are traveling to the same destination. A freight train may arrive at its destination ahead of schedule but have individual railcars that have already failed to meet their trip-plan schedule. Lastly, there may be practices where certain delays are not included in travel time.

The volume to capacity ratio measures the extent that fixed infrastructure is being used. When this value is closer to 1, the network is congested and has limited ability to handle future traffic. This particular metric is not a measurement of traffic dynamics, but rather answers the question of where capacity constraints are prevalent in the network.

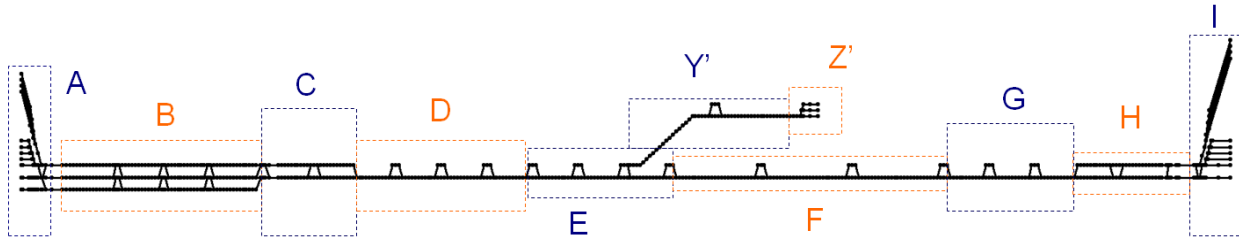
### **3.8 Spatial Analysis**

Software has been developed that graphically depicts congestion problems in real time by dynamically assigning highway links to a level of service based on GPS-enabled phones and other third-party data sets (Figure 3-9). In rush hour, there can be thousands of vehicles generating data for this service. Traffic managers can pinpoint locations of bottlenecks and congestion due to construction or accidents. Identifying important bottlenecks similar to Figure 3-9 is important for railroad planners but there are complications. The railroad generates significantly less data within the same time frame. Most links in rail networks do not exceed 30 trains per hour per direction. In a freight railroad network, flow is often measured in trains per day. Train traffic is much more discrete and not as analogous to a fluid as highway traffic. However, the small volume of rail traffic on each link allows a dispatcher to carefully monitor the performance of each train on the portion of the network under their control.



**Figure 3-9: Google Maps showing traffic congestion in Los Angeles in the evening rush hour.**

Rather than using real-time data, a railroad could use historical or simulation data to identify bottlenecks. Similar “heat” maps, similar to Figure 3-9 can be developed for railway traffic when delays are analyzed over a period of days or weeks. An important consideration is that trains usually stop at discrete locations along the line where there is an absolute signal, typically associated with a passing siding, crossover or intersection of rail lines. These absolute signals or control points can then be grouped into zones (Figure 3-10). The delay of these zones can then be compared similarly to the traffic map of Figure 3-9 except in bar chart format. In terms of bottleneck management, there are certain locations that might be better suited for a bottleneck. For example, terminals are often bottlenecks due to the amount of activities that occur there simultaneously such as car or block sorting, crew changes, and re-fueling. A terminal is designed to have multiple tracks to deal with these processes. However, if a bottleneck is on the mainline then there may not be sufficient queuing capacity to handle this traffic. This zone analysis can also measure network fluidity by measuring average speeds across a route by zone. Ideally, the average speed is constant in order to avoid a “hurry up and wait” dynamic (Williams 2011).



**Figure 3-10: Graphic illustrating how zones could be assigned to a line to analyze delays.**

Metrics are often normalized by length in both highway and railway transportation statistics. This allows for performance comparison between links in the transportation network. However, normalizing links that are orders of magnitude apart in length may lead to erroneous conclusions. For example, the value of 100 delay minutes per 100 train miles can have different interpretations. This metric easily points the analyst towards the correct conclusion when each train is 100 minutes late traversing a 100-mile segment. However, delays are not uniformly or normally distributed, so there may be analytical errors with this metric. For example, consider a zone that was two-miles long with an at-grade crossing with another railroad. 17 trains pass through the crossing with no delay and 3 trains are each delayed 13.3 minutes by cross traffic. This zone still has 100 delay minutes per 100-train miles. However, this diamond crossing would not be considered a bottleneck because 85% of the trains are not having a problem at this location. This metric applied to short segments may exaggerate the delays due to junctions and terminals. Additionally, in a zone analysis, normalization distorts the detection of cascading delays and can confound bottleneck analysis.

### **3.9 Simulation Case Study**

In this section, I will illustrate a subset of the presented capacity metrics applied to the hypothetical railroad network shown in Figure 3-10. This particular network features one major

240-mile corridor between two large terminals and a short 40-mile branch line. The western terminal receives 75% of the branch line traffic and 25% of the traffic travels to the eastern terminal. The level of initial investment in this corridor has been concentrated on the western end featuring sections of triple, double, and single track with dense siding spacing. The eastern portion of the network features sections of single track with sparse siding spacing and double track near the eastern terminal. The signaling system is representative of typical three-block centralized traffic control operation. The train type characteristics are the same as those used in several recent simulation analyses (Dingler et al. 2009, Lai & Huang 2012b, Sogin et al. 2013a). The departure times of trains are scheduled randomly from a uniform random distribution. Train performance and dispatching are simulated by the Rail Traffic Controller software that controls train movements across the network based on priority (Wilson 2012).

The capacity of this line is determined through delay-volume curve analysis for traffic levels featuring eight manifest trains serving the branch line and a variable number of bulk trains on the main route. If the MLOS is defined to be 100 minutes average delay per 100 train miles, then the corresponding delay-volume curve predicts that the maximum capacity of the line is achieved at 54 trains per day. The MLOS could also be defined by an on-time performance metric. An alternative MLOS standard could be the level of traffic where 85% of the trains arrive within three hours of the MRT over the route. This stricter standard reduces the capacity of the line from 54 trains per day to 39 trains per day.

As stated above, using a MLOS of 100 minutes average delay per 100 train miles, the capacity of this line is 54 TPD for the mixture of manifest and bulk freight traffic. The traffic on the line was then changed to a scenario of 42 TPD where eight are branch line trains, eight are intermodal, eight are passenger, and eighteen are bulk trains. The average train delay is 144

minutes for this heterogeneous traffic mixture. The bulk and branch line trains are assumed to be the base train and have a BTE of 1.0. The average BTE between the intermodal and passenger train is 1.63 as defined by Equation 3-7. The total capacity consumption of this line is then 52-base train units (BTU). Equation 3-7 uses an average delay statistic that is calculated over all train types. A disadvantage of calculating average delay this way is that the delays to different train types are expected to have different costs to various stakeholders. Additionally, Equation 3-7 does not guarantee economic pareto-efficiency where the delays to the base train are equal between all permutations that add up to the final base-train units. For example, the bulk trains are delayed 193 minutes in the heterogeneous traffic mixture that operates at 52 BTU. A homogenous line of 52 base trains per day averages 217 delay minutes. Under equal 52 BTU mixtures, the base trains are performing better in a heterogeneous mixture than in a homogenous mixture. An alternative BTE calculation is proposed in Equation 3-8b. This calculation first will determine the homogenous base-train traffic level that corresponds to the performance of the base trains in a heterogeneous mixture. The difference between these two values is then allocated to the non-base trains. Under this method, the BTE of the non-base trains is 1.51 and the BTU is 50. The average delay of the base trains is 193 minutes in a heterogeneous 50 BTU mixture as well as in a homogenous 50-BTU mixture. The difference between Equation 3-7 and Equation 3-8b are small in this case. In both methods, the BTE of the Passenger and Intermodal trains are averaged together.



$$BTE = \frac{\left(D_B + \frac{\bar{D}_M - D_B}{1 - \phi}\right)}{D_B} = 1 + \frac{\bar{D}_M - D_B}{(1 - \phi)D_B} \quad (3-7)$$

Where:

$BTE$  = Base train equivalent

$D_B$  = Average delay of base trains in a homogenous mixture of base train only at equivalent level of trains per day

$\bar{D}_M$  = Average delay over all train types in a heterogeneous mixture

$\phi$  = Percentage of base trains in a heterogeneous mixture

$$BTU = \phi V + (V - \phi V)(BTE) = \frac{1}{k} \ln \left( \frac{D_{mb}}{A} \right) \quad (3-8a)$$

$$BTE = \frac{1}{k(V - \phi V)} \ln \left( \frac{D_{mb}}{A} \right) - \frac{\phi V}{V - \phi V} \quad (3-8b)$$

Where:

$BTU$  = Base train units

$V$  = Number of trains per day

$k$  = Congestion constant from delay-volume curve of homogenous base train mixture

$A$  = Congestion constant from delay-volume curve of homogenous base train mixture

$D_{mb}$  = Average base train delay in a heterogeneous mixture

Common railway and highway performance metrics are summarized in Table 3-4. The passenger trains perform best followed by intermodal and bulk traffic. Most of the performance statistics are correlated with average train delays. However, some of the metrics will give different information. For example, the passenger train is 33-mph faster but achieves similar transportation efficiencies differing by only 5-percent. The congestion has an opportunity cost of an additional 50 miles that passenger trains could have traveled. The opportunity cost of the bulk trains was 137 miles. As indicated by the buffer-time index, 54% of the average travel time for the bulk train has to be added to the schedule in order to achieve a 95% on-time-performance. For passenger trains the buffer time index is only 23-percent. The worst bulk trains are performing poorer than the worst passenger train performers.

**Table 3-4: Performance Metrics of an Example Corridor.**

	Metric	Coal	Intermodal	Passenger
Delay (min)	Mean	193	92	41
	Min	0	0	0
	Median	158	73	34
	80th	290	140	65
	95th	495	232	96
Run Time (min)	Mean	613	414	240
	MRT	346	323	199
	Median	593	396	232
	80th	721	462	263
	95th	943	554	295
Delay per 100 Train Miles (min)		79.0	37.4	16.7
Average Velocity (mph)		24.9	43.0	76.3
95% OTP Threshold (min)		495	232	96
Distance Lost (mi)		137	70	50
Delay Percent		36%	22%	17%
Transportation Efficiency		64%	78%	83%
Travel Time Index		1.77	1.28	1.21
Buffer Time Index		54%	34%	23%
Misery Time Index		41%	30%	22%

In this simulation, the bulk and branch-line traffic constitute 61.9% of the traffic and 78.7% of the delays in the system. The OTP of these trains is sacrificed to preserve the OTP of intermodal and passenger trains. This disproportionate distribution of delay between train types is illustrated best in the single-track sections. The delays by zone are plotted in Figure 3-11. Zone F, which consists of single track with 20 miles between siding centers, generates the largest delays of any segment. This section is most likely the bottleneck of the line and causes cascading delays in the adjacent zones. The double and triple-track sections cause minimal delays to the trains. The terminals also have minimal impact on the performance of the trains. If Figure 3-11 was showing delay per 100 train miles, then the terminals would appear as the greatest source of train delay. This is not the case, since that number would be inflated by a small number of train miles in the denominator and small delays in the numerator.

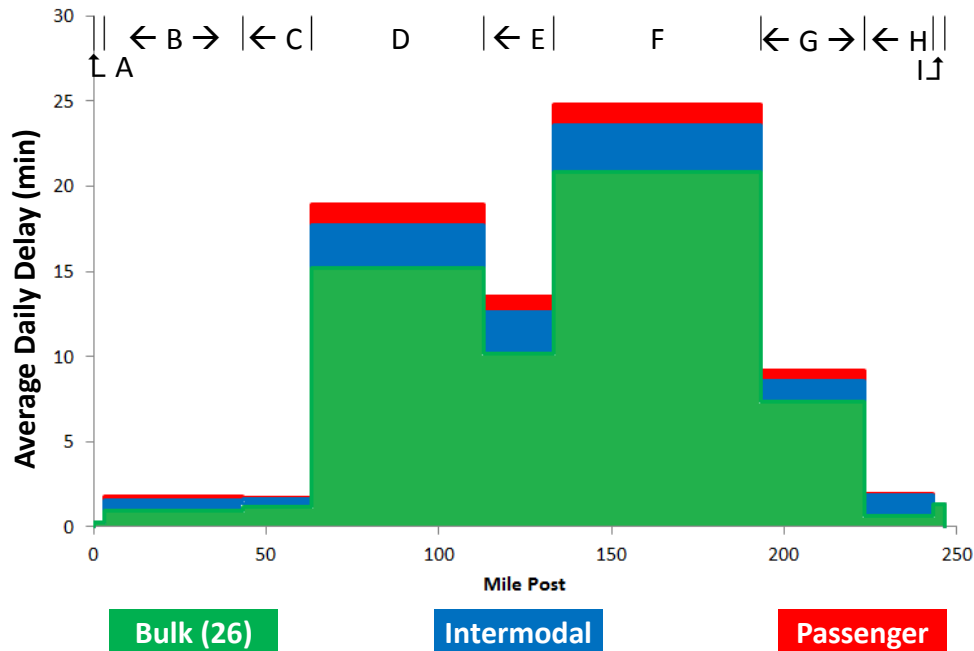


Figure 3-11: Daily train delays for zones of a railroad route grouped by train type.

### 3.10 Conclusion

Capacity can incorporate throughput, reliability, or asset utilization. It can be a subjective metric that is analyzed using several techniques. The highway mode has developed key traffic relationships that can be adapted to railroad traffic. Railroads share characteristics that are similar to fundamental highway traffic relationships, but also have unique traffic relationships that are specific to railway operating patterns, infrastructure, and control systems. Highway analysis metrics can help railroads analyze simulation data. Averages do not give a complete picture of system performance and looking at the worst performing trains can give better insight to traffic dynamics. A standardized set of railroad capacity and performance metrics may offer considerable benefit to rail traffic planners in communicating the ability of the network to support future traffic growth.

## CHAPTER 4

### RAILROAD SIMULATION TECHNIQUES FOR ACADEMIC RESEARCH

#### 4.1 Introduction

Railroad operating simulation software has developed enough sophistication to emulate actual railroad operations with reasonable accuracy. For a train to reach its destination on time, many processes need to occur without failure. Only one of those processes needs to fail to delay a train. The goal of railway simulation work in this thesis is not to build absolute predictive models that emulate railway operations. Rather, these models are designed to analyze a subset of factors that can then be compared and analyzed. Such comparative studies can help us develop fundamental relationships in railway science. The simulation software used in the studies described in this thesis is Rail Traffic Controller (RTC) developed by Eric Wilson at Berkeley Simulation. RTC is the de-facto standard for the North American railroad industry. This chapter will introduce best-practices for setting up and analyzing simulation experiments. Additionally, an overview is provided describing RTC inputs and output. This chapter will explain standards, and techniques that I have developed through my work with RTC. There may be alternative methods to achieve the same goal.

#### 4.2 Inputs to RTC

When designing an RTC experiment, it is important to understand the *.INPUT* file. This is a list of text files that contain the information such as track layout, traffic, and signals. Some of these files are much easier to manipulate than others. The most time consuming procedure of an RTC model is building a sophisticated train plan that mimics each use of the railway-track infrastructure. However, a simplified train plan using a standardized set of trains over one origin-destination pairs can improve the efficiency of the experiment. These standards reduce time

spent coding, and can increase the number of variables that can be used to analyzed. In my thesis, the most time-consuming aspect is the construction of the route in RTC. This process entails defining track arrangements, speed limits, and traffic control blocks. An experiment that changes these hard infrastructure variables will require much more manipulation within the RTC graphical user interface (GUI).

Changing the number and types of trains can be accomplished with relative ease by manipulating the *.TRAIN* file. This can be done by using a spreadsheet, script, or user created program. For every standard train, there is a base file that holds the basic information about that train including data on power, length, and weight. The script or spreadsheet requires the user to input a schedule and then the algorithm will compile the standard trains in a master train file. A script is faster than a spreadsheet and can create multiple train files from a list of inputs; however, a spreadsheet may be easier to manipulate.

When maximum train speed is analyzed, there are several important considerations. First, the speed of the train may be limited in multiple locations. Links between nodes are assigned speed limits by train type. These limits are the track speed that apply to the train as long as a portion of the train is occupying that particular link. RTC will automatically calculate the braking and acceleration curves needed for a train to meet each link's speed limits. These values are stored in the *.LINK* file. The global speed limit for a train is stored in the *.TRAIN* file within the route table. So if the link speed limit is 110 mph within the *.LINK* file and the speed limit is 79 mph in the *.TRAIN* file then the train will adhere to 79-mph maximum speed even though the track is rated higher. Another location to control the speed of the trains by type is in the *.ASPECT* file. This is more important for speed signals associated with interlockings, and not recommended as a tool to manipulate train speed for experimental purposes. In most analyses,

there will be more permutations of the *.TRAIN* file than the *.LINK* file so manipulating train speeds through the *.TRAIN* file may be preferable. Table 4-1 shows example factors that may be investigated and the corresponding locations to manipulate that variable.

**Table 4-1: Locations for Manipulating Capacity Factors within RTC.**

Capacity Factor	Recommended Location to Change
Siding Spacing	GUI
Siding Length	GUI or <i>.EQUATION</i> (if changes are small)
Siding Speed/Crossover Speed	GUI or <i>.LINK</i>
Grade	<i>.NODE</i> or <i>.EQUATION</i> (if grade length is small)
Curvature	<i>.LINK</i>
Maximum Train Speed	<i>.LINK</i> or <i>.TRAIN</i>
Number of Trains	<i>.TRAIN</i>
Train Type	<i>.TRAIN</i>
Train Length or Weight	<i>.TRAIN</i>
Amount of Double or Triple Track	GUI

### 4.3 University of Illinois at Urbana-Champaign RTC Standards

#### 4.3.1 Naming Standards

An important aspect of RTC case structure is to develop a naming convention that will yield important information about the RTC case. Coding file names will help expedite data analysis by reducing the number of lookup tables. For example, in the analysis of heterogeneity on double track, all cases were named by the number of coal, intermodal, and passenger trains present in the simulation. A case with the title “030-008-012” would mean that there were 30 coal, 8 intermodal, and 12 passenger trains present in this simulation. In the initial experimental design, there was only one maximum passenger-train speed tested so this nomenclature sufficed. However with further analysis this case code could be modified to incorporate the speed of the passenger trains. “030-008-012-110” would be a coding scheme that would incorporate

maximum passenger train speed. Another example coding scheme used was “A0-16”. Where the first character represented a corridor upgrade strategy. The second character related to the amount of double-track sections added, and the last number after the dash was the number of trains per day. “12312” was another coding scheme to correspond to a unique permutation of 5 factors that each had 3 levels (low, medium, high).

Similar to coding file names, coding train names will help analysis. Actual railroads have conventions for determining train numbers and so should the academic analyst. In this work, trains were coded in the form of “Type-Number-Direction-Day”. Within RTC, a coal train traveling westbound on the third day of the simulation would be coded as “C001W-3”. This code will provide users more information about trains in the results files without having to reference the train file to check type and direction. Regardless of the train name, RTC will automatically append the simulated day number to the end of the train.

Node numbers can also be coded to give a user more information. My work used the CSX standard “SSSMMM.MMMTT”. The first three characters (*S*) represent the route that the node is on. The next 7 characters (*M*) denote the milepost of the node between 000.000 and 999.999. The last two characters (*T*) denote the track number that the node is on. This standard will yield more precise information about a node than a simple integer. This naming convention quickly identifies, the territory, location, and track number of the node. This becomes useful when reading result files such as the *.DELAY*, *.ROUTE* and *.CONFLICT* files.

#### **4.3.2 Train Standards**

It is useful to have a standardized set of trains to create a *.TRAIN* file from. The University of Illinois at Urbana-Champaign trains used in simulation analyses are summarized in Table 4-2. The freight train characteristics were based on the Cambridge Systematics National Rail Freight

Infrastructure Capacity and Investment Study conducted for the Association of American Railroads (AAR) (Cambridge Systematics 2007). Freight car tonnages and lengths were based on averages for each car type. The power-to-weight (horsepower per trailing ton) ratios were based on expert opinion and information from the 2002 Transportation Research Board Workshop on Railroad Capacity and Corridor Planning (Galloway 2002). The passenger train was based on the existing Amtrak Cascades service in the Pacific Northwest and the proposed consist used for planning the 110 mph service between Madison and Milwaukee, Wisconsin. The passenger train stops were spaced at 32.4 mile intervals based on the current Amtrak average station spacing on routes in California, Illinois, Washington State, and Wisconsin.

**Table 4-2: Standard Trains Used in RTC Analysis.**

	Bulk	Intermodal	Passenger
Power	x3 SD70 Locomotives	x5 SD70 Locomotives	x2 P42 Locomotives
No. of Cars	115 hopper cars	75 cars	7 Horizon Cars
Length (ft.)	6,325	5,659	740
Weight (tons)	16,445	5,900	800
Max Speed (mph)	30-50	51-70	79-110
			30 miles between stops

### 4.3.3 Train Scheduling

The train schedule or timetable can influence the performance of trains across a railroad network. A “good” schedule will minimize train delays and can result in greater capacity from the network than a “poor” schedule. Some simple ways to quantify the quality of a train schedule may be to determine the number of crossing train paths on the same track or determine the frequency that a fast train is scheduled behind a slower train. Determining the quality of the train schedule before simulation is time consuming and was not a goal for this research.



Freight railroads in the United States generally use a relatively flexible scheduling system in which the initial departure time, and possibly some intermediate points along the route are loosely scheduled. However, meet and pass locations are determined in real time by dispatchers managing the traffic. These locations and the trains schedule vary widely from day to day. Martland (2009) has characterized this style of traffic management as “Improvised”. Conversely, in many European and Asian countries, much more precise scheduling is used which Martland (2009) terms as “Structured”. These two approaches have implications for how effective schedule modification will be in exploiting infrastructure capacity.

There are many methodologies on how to assign trains to slots. One approach is create the slots such that they are evenly dispersed throughout the day and alternate between eastern and western terminals of the route. For example, at 24 trains per day. The first train would be scheduled to leave the eastern terminal at 00:00 and the next train would then leave the western terminal at 01:00. A train is leaving each terminal every 2 hours. Once the slots are established, they are assigned a train type. In Dingler (2010), a worst-case scenario is assumed and used a repeating pattern. For example for a mixture of 75% coal and 25% intermodal trains, the departure sequence from a terminal would be *coal, coal, coal, intermodal, coal, coal, coal, intermodal*. The intermodal trains are not type fleeted and would likely cause an overtaking conflict with the slower coal train.

In Chapter 5, passenger trains were given preference in slots during daylight hours between 07:00 and 19:00. If the mixture of trains was still freight dominant, then the passenger trains would first strive to have service every three hours and then concentrate on the morning and evening rush hours. If the corridor was passenger dominant, then the remaining evening slots could be allocated to passenger service. Alternatively, all trains can be scheduled at  $12:00 \pm 12$

hours to remove any scheduling assumptions. RTC would create a random schedule each day with the performance of “good” and “poor” schedules averaged together.

#### **4.4 RTC Signals**

The research in Chapters 5 through 8 assumes a mainline with Centralized Traffic Control (CTC) where signals govern both train spacing and track authority. CTC enables turnouts to be lined for train routes remotely and eliminates delays from stopping a train to operate a switch manually. RTC can best model CTC and cab signal operations. RTC does not handle the nuances of dark territory controlled by Direct Traffic Control (DTC) or Track Warrant Control (TWC). RTC will capture the meet delay from manual switch operations but not track authority. RTC can be instructed to increase following distance between trains by the nodes in the dark territory. There is a node parameter that specifies a time headway constraint. Additionally, coding long absolute signals will achieve the effect of limiting distance between trains in the same directions.

If there are errors coding the infrastructure for an RTC model, they will likely be detected when signals are added. Signals need links with directions clearly defined and an “allowable path” between the start and end of the signal block. Link directions are best defined by the node mileposts such that the ascending direction corresponds to only one cardinal direction (East). A signal block must have a pathway through the network defined by the “next allowable node” dialogue boxes on the link parameters menu. These nodes specify which node a train can travel to once it departs the link it currently occupies. Fortunately, most link direction errors and allowable nodes can be fixed through the “Network→Analyze” tool. Link directions and allowable next nodes are often poorly defined near interlockings. Improperly coded switches

will interfere with automatic coding of “next allowable node”. Turnouts must have the correct normal and reverse movements coded. These errors will also arise if links are deleted.

There are two types of signals within RTC: absolute and intermediate. As in the case with real railroads, dispatchers can request changes to absolute signals, but not intermediate signals. Absolute signals grant trains authority and correspond to a safe pathway through an interlocking plant. The path of a signal block always assumes a shortest-path algorithm, however the user can override this by specifying a “via node”. Intermediate signals govern safe following distance between trains. RTC’s dispatching algorithm handles only one train per link such that an intermediate signal block should cover more than one link. This allows for two trains to occupy to be inside a permissive signal block. For RTC purposes, intermediate signals should never be coded to control diverging moves through turnouts.

#### **4.5 .INPUT Files and .OPTION Files**

Every RTC case requires an individual *.OPTION* and *.INPUT* file (Figure 4-1) with the same name. The RTC program opens the *.OPTION* file and then immediately reads the *.INPUT* file. Usually, each run in an experimental design matrix will require corresponding *.OPTION* and *.INPUT* files. Many of the RTC input files for a series of experiments will be static or change less frequently than other input files. The default in RTC is to have all input and output files in the same directory, and with each scenario having a unique set of inputs. In many cases, scenarios may use similar inputs. For example, a researcher might want to compare 20 TPD and 30 TPD on a single-track line. If a change is made to the infrastructure in the 20 TPD case, then that change must be repeated in the 30 TPD case. Fortunately, the *.INPUT* file can be manipulated so these two scenarios reference the same infrastructure inputs. The number of

inputs in an analysis should be minimized so changes made to one case updates all other cases that reference that input.

```

----- R T C 2.110      I N P U T      F I L E -----
Option: C:\RTC\DOE2\Cases\Base-Traffic.OPTION
Label: C:\RTC\DOE2\Routes\Base.LABEL
Division: C:\RTC\DOE2\Routes\Base.DIVISION
Crew: C:\RTC\DOE2\Routes\Base.CREW
Node: C:\RTC\DOE2\Routes\Base.NODE
Link: C:\RTC\DOE2\Routes\Base.LINK
Equation: C:\RTC\DOE2\Routes\Base.EQUATION
Line: C:\RTC\DOE2\Routes\Base.LINE
Signal aspect: C:\RTC\RTC.ASPECT
Signal block: C:\RTC\DOE2\Routes\Base.SIGNAL
Permit: C:\RTC\DOE2\Routes\Base.PERMIT
Form A: C:\RTC\DOE2\Routes\Base.FORM_A
Form B: C:\RTC\DOE2\Routes\Base.FORM_B
Track occupancy: C:\RTC\DOE2\Routes\Base.TOC
Locomotive: C:\RTC\RTC.LOCO
Fleet: C:\RTC\DOE2\Routes\Base.FLEET
Run times: C:\RTC\DOE2\Routes\Base.RUNTIMES
Train: C:\RTC\DOE2\Train Files\Traffic.TRAIN
Field data: C:\RTC\DOE2\Routes\Base.FIELD
Track profile: C:\RTC\DOE2\Routes\Base.PROFILE
WorkStation: C:\RTC\DOE2\Routes\Base.WORKDATION
Geography: C:\RTC\DOE2\Routes\Base.GEOGRAPHY
Environment: C:\RTC\DOE2\Routes\Base.ENVIRONMENT
Node conversion: C:\RTC\DOE2\Routes\Base.NODECONVERT

Timetable: C:\RTC\DOE2\Results\Timetable\Base-Traffic.TIMETBL
Route: C:\RTC\DOE2\Results\Route\Base-Traffic.ROUTE
Report: C:\RTC\DOE2\Results\Reports\Base-Traffic.REPORT
Summary: C:\RTC\DOE2\Results\Summary\Base-Traffic.SUMMARY
TPC: C:\RTC\DOE2\Results\TPC\Base-Traffic.TPC
History: C:\RTC\DOE2\Results\HiDTory\Base-Traffic.HISTORY
Conflict: C:\RTC\DOE2\Results\Conflict\Base-Traffic.CONFLICT
Arrival: C:\RTC\DOE2\Results\Arrival\Base-Traffic.ARRIVAL
Comparison: C:\RTC\DOE2\Results\Other\Base-Traffic.COMPARE
Noise: C:\RTC\DOE2\Results\Other\Base-Traffic.NOISE
EZone: C:\RTC\DOE2\Results\Other\Base-Traffic.EZONE
Dwell: C:\RTC\DOE2\Results\Other\Base-Traffic.DWELL
Wake: C:\RTC\DOE2\Results\Other\Base-Traffic.WAKE
Hold: C:\RTC\DOE2\Results\Other\Base-Traffic.HOLD
Delay: C:\RTC\DOE2\Results\Delay\Base-Traffic.DELAY
SCRAM: C:\RTC\DOE2\Results\Other\Base-Traffic.SCRAM
Energy: C:\RTC\DOE2\Results\Other\Base-Traffic.ENERGY
STop to STop: C:\RTC\DOE2\Cases\Base-Traffic.DTOP2DTOP
Equipment: C:\RTC\DOE2\Results\Other\Base-Traffic.EQUIPMENT
PTC: C:\RTC\DOE2\Results\Other\Base-Traffic.PTC
PTC audit: C:\RTC\DOE2\Cases\Base-Traffic.PTCAUDIT

```

Inputs

Outputs

**Figure 4-1: Example .INPUT file used to manage inputs and outputs of an RTC model.**

The .INPUT also specifies the locations of the results files. These RTC outputs could be in the same directory, however, I prefer to have an output folder for each common output file used in analysis. Each file referenced in the .INPUT file must correspond to an existing file or else RTC will crash. The specified file locations for the RTC outputs do not exist and will not

cause RTC to fail. However, if there is a typo in the location for an output file, then RTC will crash after dispatching the case. Figure 4-1 is an example *.INPUT* file used where there are both infrastructure and traffic factors manipulated. “Base” corresponds to the infrastructure inputs in the “Routes” directory that are static. “Traffic” corresponds to variables manipulated in the *.TRAIN* file. The case name and the outputs have the same unique name, “Base-Traffic”. This identifies which infrastructure configuration is being used and the traffic composition used in the scenario.

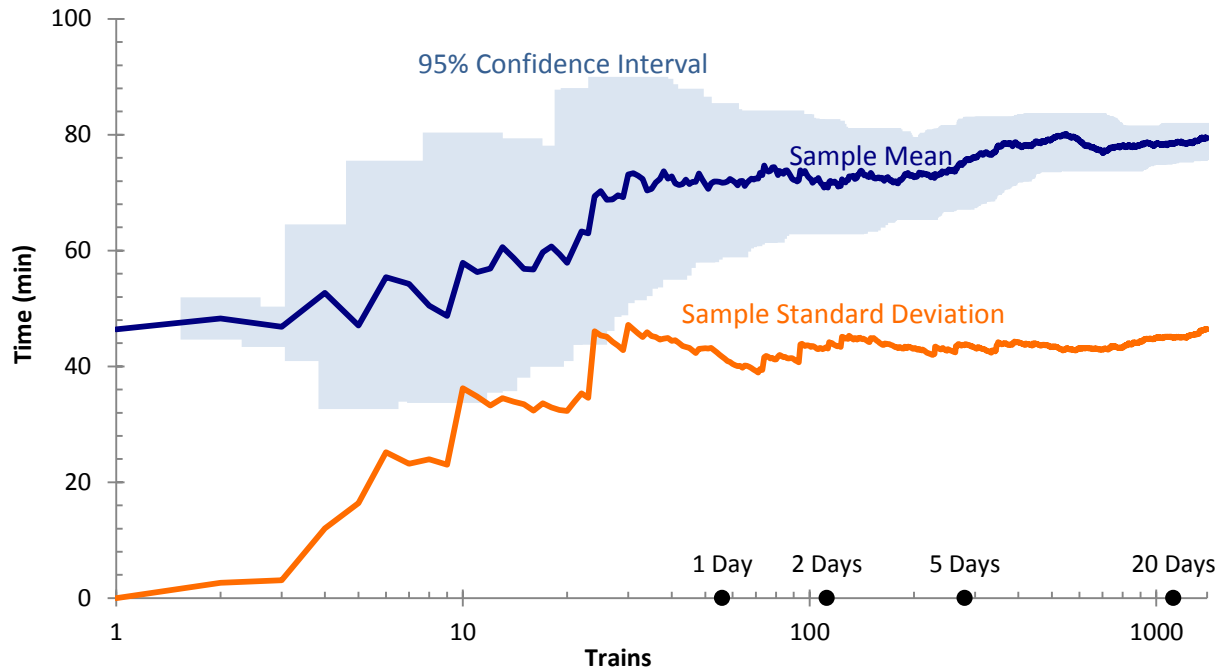
## 4.6 Data Generation

RTC has the capability to add randomization to the simulation models. This randomization is usually applied to the departure times of the trains. In this case a random departure time is assigned to a train with this feature enabled. This departure time is generated from a random uniform distribution. For a train scheduled to leave at 12:00  $\pm$  2 hours, each minute within this 4-hour interval has an equal probability for a train to depart. Randomization in the departure times serves several purposes. It can represent actual variation in departure times experienced in railroad operations. Additionally, randomization allows RTC to make different decisions for resolving similar conflicts. For example, if RTC consistently favored an eastbound train over an equal priority westbound train, then the randomization gives an opportunity to make a decision and remove this source of bias.

In order to run different replicates of the same simulation, a user needs to populate the *.SUITE* file. This can be done using a text editor or within the RTC interface. The *.SUITE* file will always be located in the folder where RTC is installed (C:\RTC). The *.SUITE* file is divided into two levels. The highest level are the “suites” that can contain up to 16 “cases”. All of the

cases within a suite must have the same file path. An individual case refers to one combination of experiment variables. The “seed” refers to the randomization. Simulating the same Seed will yield the same results while simulating different seeds should yield different results. A keystroke manipulator such as “AuotIt” can be developed to quickly edit the *.SUITE* file and run the corresponding batch within RTC. This is particularly useful if more than 192 cases (including replication) are being simulated.

A good RTC analysis will have enough days simulated to build a strong confidence interval over the average (Equation 4.1). Simulating more days will reduce the confidence interval. Figure 4-2 shows the convergence of the sample mean, sample standard deviation, and confidence interval when more trains are simulated. After one day of simulation, the sample mean and sample standard deviation were approximately equal to their final values. Simulating 19 extra days had minimal effect on these estimates; however, the extra data points created a tighter confidence interval. Simulating too many days likely yields redundant information. If traffic volume is a variable of interest, then it is expected that greater capacity utilization will yield more variability and increase the delay standard deviation (AREA 1931). So when traffic volume is low, the standard deviation will be small as well as the number of trains simulated. When traffic volume is high, it is expected to have a higher standard deviation as well as a greater number of trains simulated.



**Figure 4-2: Convergence of sample mean, sample standard deviation, and confidence of the mean (shaded area) for 56 trains per day on 73% double track.**

$$\mu = \left[ \bar{x} - t^* \frac{s}{\sqrt{N}}, \bar{x} + t^* \frac{s}{\sqrt{N}} \right] \quad (4-1)$$

Where:

$\mu$  = Confidence interval of population mean, (infinite days simulated)

$\bar{x}$  = Sample mean

$t^*$  = Upper critical value of the t-distribution with n-1 degrees of freedom at  $((1 - \alpha)/2)$

$s$  = Standard deviation of sample

$N$  = Sample size

Another important consideration is the number of days in a simulation versus the number of replicates of that particular simulation. For example 30 days of traffic could be generated by simulating six replicates of five days. Alternatively, 30 days of traffic could be generated by simulating three replicates of ten days. Each simulation requires a “warm-up” period to populate the route so the statistical trains can interact with traffic on the line. There is also a “cool-down” period so the last statistical trains can reach their destinations. So in the first example, 30 days of

traffic is being generated with 10 additional days also generated, but not used for statistical analysis. In the second example, 6 additional days are simulated, but not used in the final analysis. However, a longer simulation increases the probability of a simulation failing midway through. In a congested network, these failures usually occur due to complex train interactions, where RTC can no longer determine a feasible solution for dispatching these trains.

#### **4.7 Design of Experiments**

Design of experiments (DOE) software can systematically select factor levels that will yield tight parameter estimates in a multivariable regression analysis. This can lead to more information from less data and allow a researcher to look at more factors that may affect railway capacity. DOE is powerful in screening factors. The factor selection becomes critical in the design. A small factor range between the minimum and maximum will result in higher resolution and accuracy over that area of the theoretical response surface. A larger factor range can result in tighter confidence intervals of the parameter estimates (Sall et al. 2012). DOE can be confounding when one factor is responsible for much of the variation and this one factor may point a researcher away from other relationships that may be interesting.

#### **4.8 Data Analysis**

RTC generates numerous result files that provide different information about the performance of the simulated trains. Depending on the experiment, anywhere from 100 to 100,000 result files may be generated. Some software tools to external RTC are needed to read the result files and compile the data for further analysis. The next section will explain the type of information in each output file.



#### **4.8.1 External Software to RTC**

##### *Notepad++*

This is a text editor that can open RTC input and output files. This software can vertically type or select text by using the short cut “[Shift]+[Alt]+[up or down]”. This software can open, save, and close multiple files simultaneously. Additionally, the software can do “find & replace” over all opened files. Notepad++ is useful for creating, editing, and managing RTC files.

##### *AutoIt*

AutoIt simulates keyboard and mouse inputs into the computer. This has a simple coding language and can be learned quickly. AutoIt has been used to automate repetitive tasks such as editing *.SUITE* files and running the corresponding batch in RTC (see pg. 61)

##### *JMP*

*JMP* is a commercial statistical analysis package. The software has powerful data visualization tools that can quickly graph data. The software has linear and non-linear regression procedures as well as neural network modeling capabilities. *JMP* can also be used to create experimental designs and fit probability distributions to multiple datasets simultaneously.

##### *SAS*

SAS is another commercial statistical analysis package with similar capabilities to *JMP*. SAS is more customizable than *JMP* and was also used in this work.

## 4.8.2 RTC Output Files

### *.TIMETBL*

Generates a table of train departure times by node location. These nodes must be specified ahead of time by checking the corresponding box in the node dialog box in the user interface of RTC. This file is not created until a user brings up the Timetable Screen within RTC.

### *.ROUTE*

This is a spreadsheet showing train paths by node. This file will provide the necessary information to determine a trajectory of a train. Each row represents a node along a train-path. The milepost of that node is reported along with instantaneous velocity, time the head-end of train arrived at the node, the time it departed the node, and the time the tail-end of the train departed that node. There is some additional information here that can be useful. Stop delay and switch delay are reported at these nodes. Lastly, the signal aspects that the locomotive engineer sees are recorded at the nodes. A stop signal is not reported because the subsequent clear signal will overwrite it. The *.ROUTE* can be saved as a .csv file or a text file as a checkbox option within the GUI.

### *.REPORT*

There are multiple sections of data to this file:

1. Performance of individual trains across their entire routes. Key statistics reported are run time, true delay, average velocity, and fuel consumption. Most of the analysis in this work used the individual train data reported here.
2. Performance of individual trains by subdivision. Key statistics reported here are run time, delay, and average velocity across the subdivision.

3. This section includes statistics on the links of the network. It will report train counts over links in the network as well as how long trains occupy that link. If a study wanted to analyze the usefulness of various crossover configurations, this would be the appropriate place to start.
4. Stop delays by train type at the nodes in the network.

### *.SUMMARY*

This file will report similar performance statistics as the first two sections of the *.REPORT* file: run time, average velocity, delay, etc. This will aggregate the train performance by train type and report the corresponding average. On-time-performance by train type is also found in this file.

### *.TPC*

This is the output file from the internal Train Performance Calculator within RTC. This file is only created when a user opens the TPC screen within RTC. Each time a user clicks on a different train, a new TPC run is appended to the of the end of the *.TPC* file. Similar to the *.ROUTE* file, the data for distance, time, and velocity are reported in this file. However, this is done by TPC increment and not by node. Hence, the *.TPC* file will give more accurate information on where a train is located. The *.TPC* file will report information not found in the *.ROUTE* file such as the throttle position of the locomotive and cumulative fuel consumption.

### *.HISTORY*

This output file is a record of all the decisions RTC made while dispatching the simulation. If a case has already been dispatched, then RTC can load the *.HISTORY* file instead of re-computing a dispatching solution. The *.HISTORY* file is the largest RTC output file and can often be over 1

GB of hard drive space. I recommend compressing the *.HISTORY* files into a .zip file to save hard drive space.

### *.CONFLICT*

This file is a record of the conflicting train paths and how RTC resolved that particular conflict. Additionally, crew expirations, and delayed departure time will be reported in this file. These conflicts are the primary delays and may not be the root cause of why a particular train is being delayed. The best way understand the contents of this file is to follow an RTC animation of what occurred in the time frame. If the user of the *.CONFLICT* file does not have access to RTC then a .PDF of the time-distance chart can help. In general, a signal meet lookout is a meet between two trains, and the lower priority train will diverge and take the siding. A pass restriction conflict is between two trains in the same direction. This may be an overtake, a yard conflict, or a train trailing behind a slower train.

### *.ARRIVAL*

This file will display when a statistical train arrives and departs the event nodes specified in the *.TRAIN* file.

### *.DWELL*

This file is a record of the dwell times. The dwell time at a particular node is reported by train. This is a location to verify the alternate node logic.

### *.HOLD*

This is a record of the dispatching orders given to the simulated locomotive engineers. These data are in a sentence format, not a spreadsheet. If standardized train names, node numbers are implemented, then the information presented in this file will be self-explanatory.

### *.DELAY*

There are multiple sections to this file.

1. The first section is a record of decisions made in order that a train meets its scheduled arrival times at event nodes.
2. This is a record of trains that violate the maximum hours of service threshold that a crew can be on duty. This is listed by train, location, and time.
3. This is similar to the information found in the *.CONFLICT* file. Each conflict is listed and notes the delayed train and the train receiving higher priority. The node where the delayed train will incur its stop delay is also listed. This will not show delays that do not include a stopping event such as a rolling meet.
4. This shows the cumulative stop delay by node.
5. This will show all the stop delays by statistical train.

### **4.8.3 Aggregating Train Data**

The next challenge after generating all the output files is to transfer the relevant information into a new database. There are multiple ways of creating a database. If the experimental design matrix is small, then manually extracting the relevant information from the *.REPORT* or *.SUMMARY* files may be feasible. However, if many simulations are run, then the process needs to be automated. One method is to use an external program that combines all the *.REPORT* files in a folder into one large text file. This large text file is then inserted into Excel where a series of “If” functions and text functions pull the actual information into a spreadsheet format. Another

method is to write a script to read certain *.REPORT* files and store the relevant information in RAM. Some RTC files have multiple sections containing different information, so the scripts or Excel formulae need to be able to determine where to pull data. Once each report file is read, then the program will create a spreadsheet designed to the user's specification. Preliminary calculations can be done here. Once a base script is written, it is easily adaptable to other situations. An ACCESS database can also be developed to read RTC output files. An example PYTHON script that reads run time, distance traveled, and delay from the report file is below.

```
#-----Objects Used in Script-----
class train():
#Train Object. Contains important performance data about each train
    def __init__(self):
        self.train_name = 'X###N-#'
        self.train_type = "T"
        self.direction = "N"
        self.distance = 0.0
        self.avg_spd = 0.0
        self.run_time = 0.0
        self.total_delay = 0.0
        self.term_dwell = 0.0
        self.day = 0

#-----Functions Used in Script-----
def write_train(T,sim_name):
#Writes a line of data for one train
#Inputs: #T = Train object, #sim_name = Simulation name
#Outputs: one line of data

    Run_Number = sim_name[len(sim_name)-3:]
    items = [sim_name,Run_Number,T.name,
            T.type, T.run_time/60, T.total_delay/60]
    string = ""
    for col in items:
        string = string + "," + str(col)
    return string[1:]+ "\n"

def read_all(all_trains,simulation):
    filename = simulation + ".REPORT" #Name of report file
    report_file = open(filename,'r') #Open the report file
    Train_List = {} #Dictionary of trains, called by:
                    # (Train Name, Simulation):
    counter = -1
    max_days_simulated = 0
    incomplete = -1
    for line in report_file:
```

```

#Search for incomplete simulation
inc = line.find("DISPATCH NOT COMPLETE")
if inc <> -1:
    incomplete = 1

#Search for the overall performance data
a = line.find("Statistics for included run-time trains")
if a <> -1:
    counter = 0
    if counter >= 0:
        counter = counter + 1
        if len(line)>=350:
            if line[83]=="%" and line[170]=="%":
                #Train Statistics
                New_Train = train()
                New_Train.name      = line[7:14]
                New_Train.type      = line[7]
                New_Train.direct    = line[11]
                New_Train.avgspd    = float(line[30:36])
                New_Train.run_time  = float(line[69:75])
                New_Train.total_delay = float(line[155:160])
                New_Train.dwell     = float(line[93:97])
                New_Train.day       = int(New_Train.name[6:])
                Train_List[line[7:14]] = New_Train
                max_days_simulated = max(max_days_simulated, New_Train.day)
    if counter >=0 and line.find("Case")<>-1:
        counter = -1
report_file.close()

#Checks to see that there is least 2 full days simulated
if incomplete<>-1:
    Last_day = max_days_simulated - 2
    I = "I-"
else:
    Last_day = max_days_simulated
    I = ""
for t in Train_List:
    day = int(t[6:])
    if day<=Last_day:
        all_trains[t,I+simulation] = Train_List[t]
return all_trains

#-----Main Program-----
print "Name of text file with list of cases to read:"
case_list = str(raw_input())
case_file = open(case_list+".txt",'r')
all_trains = {}
for case in case_file:
    for r in range(50,1000,50):
        sim = str(case)[:len(case)-1]+"-R"+str(r)
        try:
            open(sim+".REPORT",'r')
            all_trains = read_all(all_trains,sim)
            print "Read",sim
        except IOError:
            z ='do nothing'

```

```

#Write new .csv file
Header_Row = "Simulation, Run_No, Train Name, Train Type, Run Time, Delay \n"
ofile = open('resultss.csv','wb')
ofile.write(Header_Row)
for entry in all_trains:
    ofile.write( write_train(all_trains[entry],entry[1]) )
ofile.close()
print "Done!"
#-----End of Program-----
Comment
Function
Built in function
Value
String
Constructor

```

When a RTC fails to complete a simulation, the first line of each output file will read “DISPATCH NOT COMPLETE. RESULTS INVALID”. There are three options for how to work with invalid RTC data. 1) The data can simply be discarded. 2) The data for this failed simulation can be estimated. 3) RTC may have failed late enough in the simulation process so there are complete train paths generated independent of the problem that caused the failure in the first place. The performance of the trains that ran 24-48 hours before the point of failure can be extracted from the *.REPORT* file. The number of trains kept should still reflect the original traffic composition.

#### 4.8.4 Separating Data

The experiment database should contain data for all the trains in each simulation. The next challenge is to organize the data such that trends can be determined. In most situations, the performance of individual trains across the networks is expected to be highly variable. For the same set of parameters, it is not uncommon to have standard deviation of train delay exceeding 20 minutes per 100 train-miles. Reporting the performance of each individual train is impractical. Average train delay is often used; this should be treated as an average and not a



measure of central tendency. Delay distributions are often skewed to the right such that the average will no longer be a measure of central tendency. One should be aware that reporting averages may hide the performance of the worse performing trains which may be violating the on-time performance threshold or experiencing crew expirations.

Pivot Tables are a useful tool to calculate summary statistics over various categories. This allows for quick exploratory analysis of the data. I often setup traffic volume as a row label, train type as a column label, and the transportation statistics as the value fields. Pivot Tables are hierarchical such that multiple categories can be grouped together and sub-totaled.

#### **4.8.5 Multivariate Regression**

Regression modeling is a statistical technique that can be used over series of independent variables ( $x$ ) and one response variable ( $y$ ). Common applications include interpolation, extrapolation, and factor screening. Regression can smooth noise from a dataset and interpolate values in between the data points used to create the regression. This helps researchers determine relationships from the factors of interest. Factor screening analyzes many variables simultaneously and the variables are then ranked based upon their ability to change the response variable. Extrapolation is when a regression relationship is used to make a prediction outside of the data range used in fitting the model. This work used factor screening techniques and interpolation to derive key relationships between factors that can affect railway capacity.

In simple linear regression, an equation for a line is derived from  $x$ - $y$  coordinates where  $y$  is assumed to be a function of  $x$  (Equation 4-2). For any given  $x$ , there can only be one  $y$ -value.  $\beta_1$  is the “main effect” variable. This value corresponds to the marginal change in the response variable when the independent variable changes. In multivariate regression,  $y$  is a function of many  $x$ -variables with many different parameters. Large parameter estimates of individual

factors correspond to large marginal changes in the response variable. However, there could be a unit bias in the parameter estimates and comparison would require a normalization between factors. Arc Elasticity or the t-statistic can be used to accomplish this. Regression software will often yield 95% confidence intervals for the parameter estimates. This interval should be entirely positive or negative and not include 0. Some important outputs from a multivariate regression model are summarized in Table 4-3 through Table 4-5

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (4-2)$$

Where:

$i$  = Index of datapoints

$Y_i$  = Responding variable of datapoint  $i$

$X_i$  = Independent variable of datapoint  $i$

$\varepsilon_i$  = Error term

The  $R^2$  term measures how much of the random variation in the data is explained by factors of in the model. A value of 1.0 means that the model explains all variation and a value of 0 indicates that the model is performing no better than the mean of the dataset. This number is influenced by the number of the variables in the model and the type of data used. In most cases, there will be replication in the experimental design such that achieving an  $R^2$  of 1.0 is near-impossible. Root mean square error is another important measure because it will provide insight into the “average” error term in the data set (Trabke4-3). Analysis of Variance (ANOVA) tables are a standard form of measuring significance between the means of groups (Table 4-4), and is another area of model diagnostics.

**Table 4-3: Fit Statistics of a multivariable regression model.**

Fit Statistics	Value
RSquare	0.5419
RSquare Adj	0.5416
Max RSquare	0.5519
Root Mean Square Error	55.39224
Mean of Response	75.77093
Observations (or Sum Wgts)	34752

**Table 4-4: ANOVA Table.**

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	22	126,053,080	5729685	1867.381
Error	34729	106,558,999	3068	Prob > F
C. Total	34751	232,612,079		<.0001*

An important consideration is which response variable,  $y$ , to use from the simulation results. This is usually run-time, average velocity, or delay. Each of these three are interrelated and have their strengths and weaknesses as a performance metric. Delay is a good measurement of congestion; however, it is difficult to compare against other train types with different MRTs. Run time will be sensitive to maximum train speeds and may not be able to distinguish between slower speeds and congestion within the network. Once a metric is chosen, then a decision about the aggregator must be made. The performance of individual trains can be left un-consolidated, grouped by the simulation replicate, or grouped by their permutation of inputs. Regressing on individual train delays will lead to a poor  $R^2$ . There is large variability in the delays of individual trains but not necessarily in the average of those trains. By having a large dataset, there should be tight confidence intervals around the mean prediction error and parameter estimates. Alternatively, the response variable can be the average train delay of each replicate. This method will have lower error terms, and a higher  $R^2$  value. The variability of a sample mean is considerably lower than the variability of an individual train delay. Lastly, the

replicates can be combined into one average train delay calculation. An important consideration is that least-squares regression can bend the response surface to areas of greater variability or where there are more data. Weighting the sample means by sample size or standard deviation within the sample may be necessary in order to have an un-biased response surface model.

**Table 4-5: Effect Tests of Variables in Model.**

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Allocation	2	2	346250	56.4238	<.0001*
Double Track(0.19,1)	1	1	21415513	6979.601	<.0001*
Volume(12,60)	1	1	39939860	13016.93	<.0001*
Hetero(0.16667,0.83333)	1	1	547058	178.2936	<.0001*
Pass Speed(40,110)	1	1	427522	139.3351	<.0001*
Allocation*Double Track	2	2	251654	41.0087	<.0001*
Allocation*Volume	2	2	885404	144.2824	<.0001*
Double Track*Volume	1	1	33781448	11009.82	<.0001*
Volume*Volume	1	1	7098936	2313.638	<.0001*
Allocation*Hetero	2	2	129443	21.0936	<.0001*
Volume*Hetero	1	1	500970	163.2728	<.0001*
Hetero*Hetero	1	1	97291	31.7083	<.0001*
Allocation*Pass Speed	2	2	423461	69.0058	<.0001*
Double Track*Pass Speed	1	1	60996	19.8795	<.0001*
Hetero*Pass Speed	1	1	20458	6.6677	0.0098*
Pass Speed*Pass Speed	1	1	38030	12.3945	0.0004*
Volume*Volume*Double Track	1	1	3656248	1191.620	<.0001*

## CHAPTER 5

### COMPARISON OF THE CAPACITY OF SINGLE AND DOUBLE-TRACK RAIL LINES USING SIMULATION ANALYSES

*Sogin, Samuel L, Yung-Cheng Rex Lai, C Tyler Dick, and Christopher P L Barkan. 2013. "Comparison of the Capacity of Single and Double Track Rail Lines Using Simulation Analyses." Transportation Research Record: Journal of the Transportation Research Board. Washington, D.C. In Press*

#### 5.1 Introduction

Long-term freight demand is projected to increase substantially, and new passenger services are being proposed to operate over portions of the freight infrastructure (AASHTO 2007). Passenger and freight trains have different characteristics in terms of power, weight, braking performance, top speed, priority and on-time arrival performance. The unique characteristics of each train type places different demands on railroad infrastructure. Operating multiple train types on one line can introduce higher delays than operating a single train type. U.S. freight railroads traditionally host passenger trains operating at a maximum speed of 79 mph. Speeds faster than this may introduce new challenges to managing the existing capacity of a railroad. The objective of this chapter is to compare the simultaneous operations of passenger and freight trains in single and double track configurations taking into account increased maximum passenger speeds. The simulation software called Rail Traffic Controller (RTC) was used to evaluate effects of homogeneous and heterogeneous operations (Wilson 2012).

Single-track line capacity is limited by the need for trains to decelerate, stop and accelerate out of sidings to allow other trains to use the intermediate single-track sections. The largest cause of train-interference delays on single track is the meet delays at these sidings (Dingler et al. 2010). Double track configurations can mitigate this dynamic and allow the line to

operate at a significantly higher capacity. Because of these inherent efficiencies, double track lines can run more trains at higher average speeds than a single track configuration. In the absence of meets, passing conflicts and train spacing become key capacity constraints for a double track line.

For both single and double track configurations, previous research has determined that simultaneous operation of different train types consumes more capacity than homogenous operations. Vromans et al. (2006) used simulation to investigate options to improve passenger operations. Leilich (1998) and Dingler et al. (2009) used simulation analysis to analyze the delay caused by the interaction of unit trains and intermodal trains and found a capacity loss due to heterogeneous operations. Harrod (2009) used integer programming to show that it can be feasible to run higher speed trains in a single track freight network provided there is a lane available. Peterson et al. (1987) used simulation analysis of where to locate sidings in single track to accommodate higher speed trains. Sogin et al. (2011, 2012b) used simulation analyses to show there is a larger increase in delay to freight trains by adding a passenger train instead of a freight train in both single and double track.

## **5.2 Methodology**

Four key factors (Table 5-1) were considered in the simulation experiments. The different permutations of Table 5-1 can represent various shared corridor conditions. Traffic volume is defined as the total number of trains per day (TPD). Traffic mixture (heterogeneity) is the percentage of these that are freight trains and describes the train-type heterogeneity of the corridor. The parameters, total trains per day and percent freight are also described interchangeably by number of passenger trains and number of freight trains per day. The

subsequent analyses use both pairs of parameters. The maximum speed of the passenger train was analyzed at 79 and 110 mph with intermediate speeds used in a correlation analysis. The fourth factor was the number of main tracks on the line.

**Table 5-1: Experiment Factors.**

Factor	Levels
Number of Tracks	1,2
Traffic Volume (trains per day)*	0 – 80 (increments of 4)
Maximum Passenger Train Speed (mph)	79, 110
Traffic mixture (% freight trains)*	0% – 100% (Even increments of 4 trains)
*Traffic Volume and Traffic Mixture Can be also be replaced by two parameters: number of freight and number of passenger trains per day	

The full factorial of Table 5-1 was simulated. Each simulation run featured a unique combination of the four factors described in Table 5-1. Rail Traffic Controller (RTC) was used to simulate a dispatcher making decisions regarding the train movements across a particular line. RTC is commonly used by railroad capacity planners in North America to simulate train traffic. Each simulation run produced the performance of trains over a three-day period. This simulation was then repeated four times to yield 12 days of total simulated traffic. The replication allowed the dispatching algorithm to make different decisions for any given traffic mixture. For example, one simulation has 40 freight and 24 passenger trains that operate at a maximum speed of 79 mph. This simulation was repeated four times to yield performance metrics for 768 trains. The single track configuration was expected to have a lower capacity than the double track route and will not be able to operate as efficiently at higher traffic volumes.

The routes used in the analysis represent idealized freight lines. The line was simplified to facilitate comparisons between the key variables described above. The single track and double

track lines are described in Table 5-2. Each route featured one origin-destination pair with 0% grade and zero degrees of curvature. Additionally, all the trains on the line were one of the train types identified in Table 5-3. The single-track line featured passing sidings every 15 miles between siding centers where trains can meet and overtake each other. The double track line featured 45 mph universal crossovers every 15 miles where trains can change tracks. Because the double-track route has higher capacity, it will operate with less delay than the single-track route. The routes were compared to each other at traffic levels where both routes are considered congested. Traffic mixtures between 0%-25% freight trains are considered passenger-dominated corridors such as the Amtrak Northeast Corridor. Mixtures between 75%-100% freight trains are freight-dominated corridors such as the Amtrak Cascades Corridor. The single track is considered congested at 36 trains per day and the double track route is considered congested at 64 trains per day.

**Table 5-2: Route Characteristics.**

Parameter	Single Track	Double Track
Length	265 miles	265 miles
Interlocking Spacing	15 miles between sidings	15 miles between universal crossovers
Siding Length	7,920 feet	n/a
Diverging Route Speed	45 mph	45 mph
Traffic control system	2-Block, 3-Aspect Bi-directional CTC	
Average signal spacing	2.2 miles between intermediate signals	

Train slots were scheduled evenly throughout the day. So for 24 TPD, there would be a train leaving each of the two terminals every two hours. The departure times between the eastern and western terminal would be offset by 1 hour in this case. Passenger trains were given preference to operate in the slots between 6am and 8pm to reflect the short-haul daytime passenger schedules common to most existing and planned North American shared corridors. In high percent freight corridors, the slots between 6am and 8pm were allocated to both passenger



and freight trains. In lower percentage freight mixtures, the schedule was similar to temporal separation; passenger trains operated during the day and freight at night.

Individual train types vary in length, power, and weight. Each train in the simulation was based on the characteristics specified in Table 5-3. While there are many different classes of freight trains, a single type was used to facilitate comparison of the key variables in Table 5-1. The freight train characteristics were based on the Cambridge Systematics National Rail Freight Infrastructure Capacity and Investment Study conducted for the Association of American Railroads (AAR) (Cambridge Systematics 2007). Freight car tonnages and lengths were based on averages for each car type. The power-to-weight (horsepower per trailing ton) ratios were based on expert opinion and information from the 2002 Transportation Research Board Workshop on Railroad Capacity and Corridor Planning (Galloway 2002). Freight train departure times deviated from their prescribed departure times by  $\pm 20$  minutes in a random-uniform-distribution.

**Table 5-3: Train Characteristics.**

	Unit Freight Train	Passenger Train
Locomotives	x3 SD70	x2 P42
No. of Cars	115 hopper cars	11 Articulated Talgo Cars
Length (ft.)	6,325	500
Weight (tons)	16,445	500
HP/TT	0.78	15.4
Max Speed (MPH)	50	50-110
Minimum Run Time (min)	355	356-196
Unique Characteristics	$\pm 20$ minutes departure time	32.4 miles between stops

The passenger train was based on the existing Amtrak Cascades service in the Pacific Northwest and the proposed consist used for planning the 110 mph service between Madison and Milwaukee, Wisconsin. The passenger train stops were spaced at 32.4 mile intervals based on the

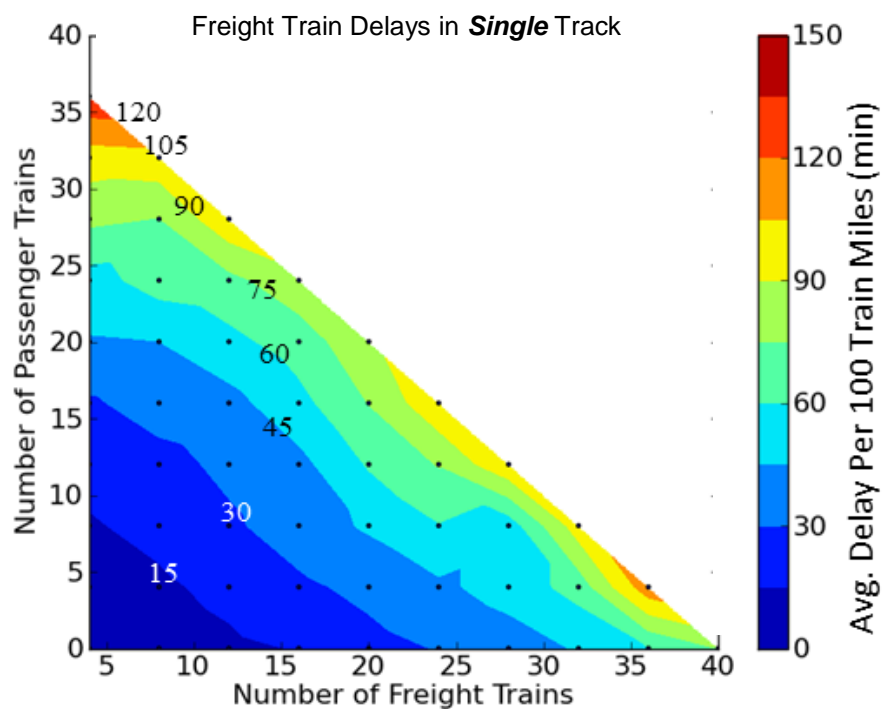
current Amtrak average station spacing on routes in California, Illinois, Washington State, and Wisconsin. The maximum speeds used were 79 mph and 110 mph. By Federal Railroad Administration regulations, train speeds are limited to a maximum speed of 79 mph without advanced signaling and highway-rail grade crossing technologies. Illinois and Michigan have, or soon will, increase the track speed to 110 mph in certain locations. All passenger trains have higher dispatching priority than freight trains (Sogin et al. 2011, 2012b, Dingler et al. 2013).

Freight train performance was measured by delay minutes per 100 train miles. Delay is defined as the difference between minimum run time and the actual run time of a particular train. The maximum speed of the freight trains was held constant in all simulations at 50 mph. Passenger train speed is an experimental design variable that varies so the minimum run times also change. This means that the delay to a 110 mph train is different from the delay to a 79-mph train. A minute of delay to a 110 mph passenger train has a higher distance cost than a minute of delay to a 79-mph passenger train. A 5-minute-stop delay for a 79-mph train costs 6.6 miles of travel, and a 5-minute-stop delay for a 110 mph train costs 9.2 miles of travel. Because of this difference in distance cost, passenger train performance was evaluated by the comparison of actual run times normalized by 100 route miles. Trains with higher maximum speeds will tend to have faster travel times. The variation in the time to travel 100 miles is also of interest in evaluating the reliability of the passenger trains.

### **5.3 Analysis**

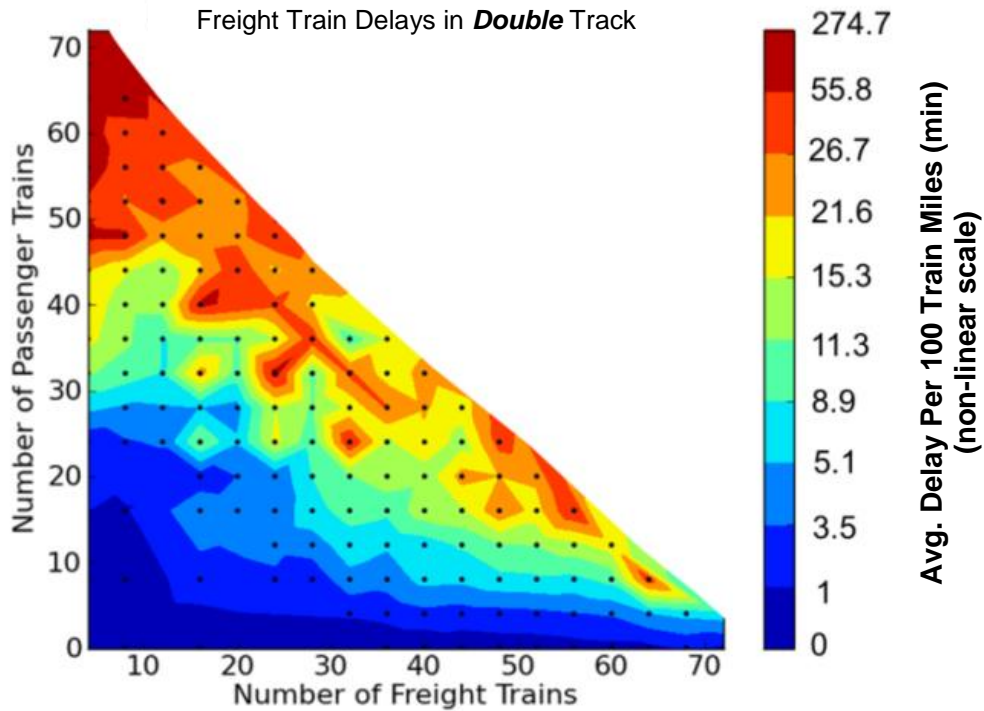
The delay-response surface for single track is shown in Figure 5-1a. The color intensity represents the average freight train delay per 100 train miles for various combinations of 110 mph passenger and 50 mph freight trains. The contours are generated from a 4-train-unit grid

based on all combinations of 110 mph passenger trains and freight trains. The contours are in increments of 15 minutes per 100 train miles. The linear density of these contours is sparse at low traffic volumes and high at increased traffic volumes. This is consistent with delay-volume curves used to describe train delay (Krueger 1999, Abril et al. 2008, Pachi 2009, Schlake et al. 2011, Lai & Huang 2012b, Lai et al. 2012). A combination of 4 freight and 24 passenger trains per day has similar freight-train delay to that of a homogenous freight line with 36 trains per day. Assuming pareto-efficiency with regards to freight train delay, this mixed operation translates into a capacity loss of 8 trains per day (from 36 to 28 trains per day). The contours are linear with an average slope less than 1 indicating that there is a consistent capacity loss due to heterogeneous operations.



(A)

Figure 5-1 cont. on next page



(B)

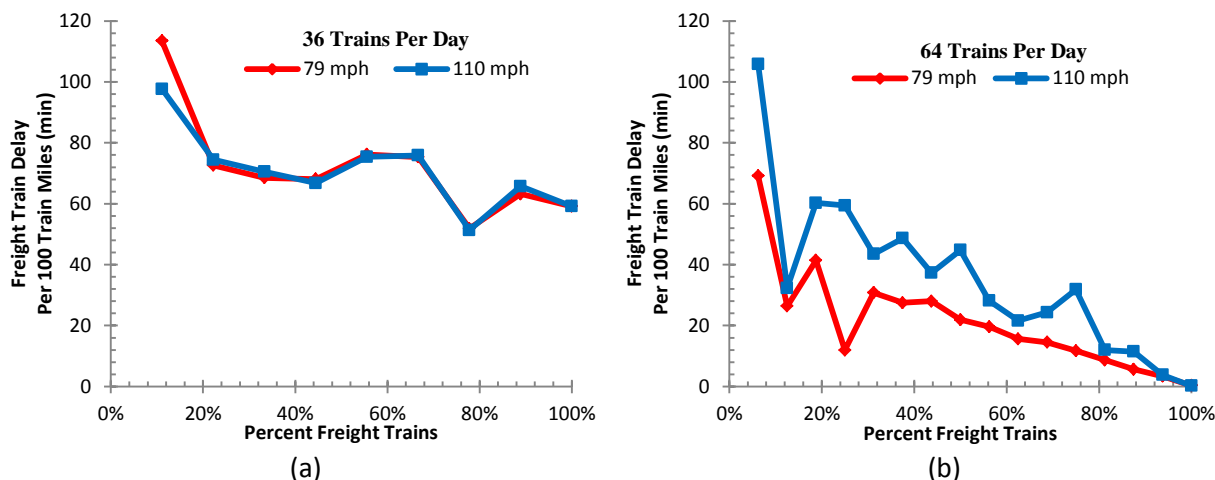
Figure 5-1: Freight train delay surface at various traffic mixtures of 110 mph passenger and 50 mph freight trains in (a) single track and (b) double track configurations. Delay is measured as average delay minutes per 100 train miles. (x-axis starts at 4 TPD and the scale is different in (b) to show greater contour definition at low levels of traffic ).

The delay-response surface for double track is shown in Figure 5-1b (note that the scale is different from Figure 5-1a). Unlike single track, the contours are non-linear. These contours are concave indicating a significant increase in train delays as the line transitions from a freight-dominated network to a passenger-dominated network. As the line approaches capacity, the contours are no longer parallel. This indicates that there are cases with higher traffic levels and lower freight train delays. In a capacity constrained network, the delays may be sensitive to the management of the train interactions once traffic levels are high as well as the traffic density. The noise in the contours is probably due to the scheduling assumptions of the trains; certain schedules have different frequencies for how often a passenger train is scheduled after a freight.

When this occurs, a passenger train is likely to catch up to a freight train and cause an overtake conflict that may delay trains traveling in the opposite direction.

On single track, much of the delay is attributed to meets with other trains (Dingler 2010). In the homogenous case, an individual freight train can be expected to be favored 50% of the time in meets with opposing trains. In heterogeneous traffic mixtures, priority will influence which trains will be favored in a meet. In a 50:50 heterogeneous traffic mixture, the freight trains may only be favored for 25% of the conflicts and take the siding in 75% of the meets. The higher priority passenger train causes this impact to the freight trains and it occurs regardless of the maximum speed of the passenger train.

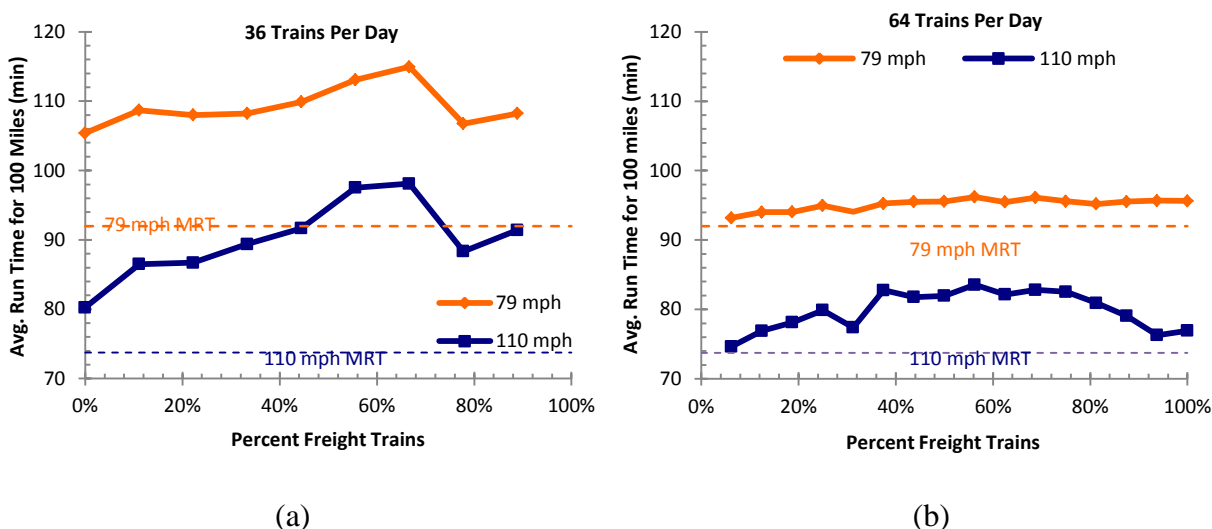
trains than 79-mph passenger trains because of the greater speed differential.



**Figure 5-2: Freight train delays in different traffic mixtures delay and different speed of passenger trains in (a) single track and (b) double track.**

Figure 5-2a shows minor differences in freight delay between 79- and 110-mph passenger trains in a congested network at various traffic mixtures. In single track and double track scenarios, freight trains perform best in freight dominated networks and experience substantial delays in a passenger dominated network (Figure 5-2a). This trend is more pronounced on

double track than on single track because removing the passenger trains removes the major source of train interference since trains no longer have to overtake each other (Figure 5-2b). In double track, 110-mph passenger trains cause more interference to the freight.

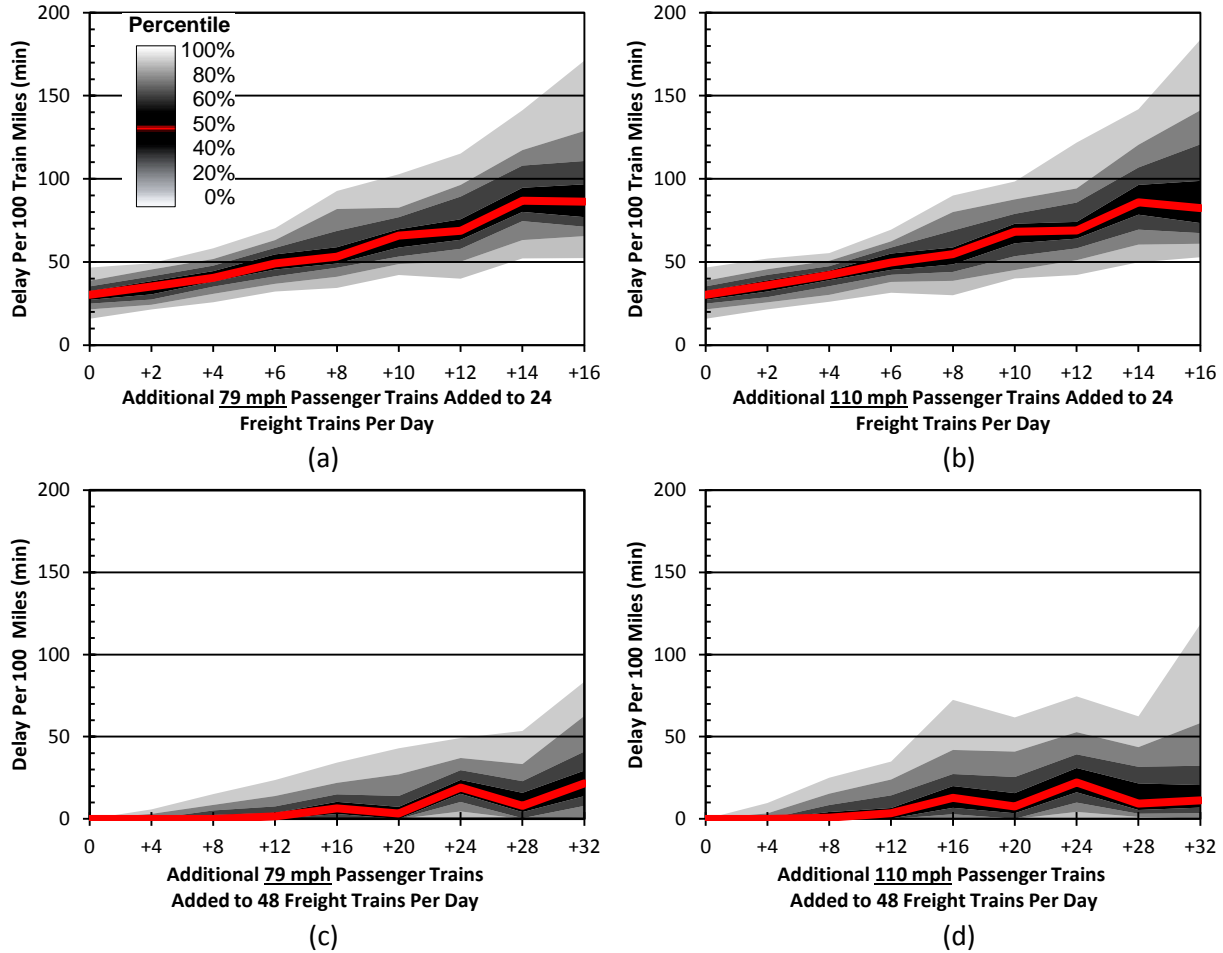


**Figure 5-3: Passenger train run time in (a) single track and (b) double track at different speeds and in different traffic mixtures.**

An important goal in passenger operations is to have run times very close to the minimum possible most, if not all the time. The average run times of passenger trains are plotted for single track (Figure 5-3a) and for double track (Figure 5-3b). The distance between the minimum run time and average run time is the delay. In single track configuration, delays for 79 and 110 mph passenger trains are similar. Both trends show the passenger trains performing best in passenger dominated networks in the range of 0-45% freight trains. The passenger trains show the longest run times in the range of 45-75% freight trains. Within this range, there are delays due to trailing behind freight trains as well as stopping for meets at sidings for other high priority passenger trains to go by. In the range of 60-100% freight trains, the run times of the passenger trains start to decrease slightly because there is less heterogeneity and fewer higher priority trains to conflict with the passenger train.

On double track, the 110-mph passenger trains are delayed more than the 79-mph passenger trains. Due to a greater speed differential, a 110-mph passenger train is more likely to catch up to a 50 mph freight train than a 79-mph passenger train. Additionally, a 110 mph passenger will also lose more time when trailing behind a freight train compared to a 79-mph passenger train. With a lower speed differential, the 79-mph passenger trains show little sensitivity to operating in heterogeneous conditions on double track. Similar to the trend on single track, the 110 mph passenger trains on double track experience higher run times in heterogeneous conditions between 45% to 75% freight trains. This may be caused by the combination of frequent passing conflicts with freight trains and the greater likelihood that an opposing train is a high priority passenger train.

Another implication of running passenger trains on the freight network is the increase in variation introduced into the freight network. Figure 5-4a and Figure 5-4b show similar distributions of freight-train delay in 10% bands as passenger trains are added to a single-track freight operation. The more passenger trains operated, the greater the variation in freight-train delay, and the more skewed the distribution will be. The performance of the worst 10% of freight trains is particularly important because these trains are the most likely to exceed a train crew's hours-of-service limit. Federal law prohibits train crews from being on duty for more than 12 hours before a relief crew must take over (Federal Railroad Administration 2013b). Therefore, increased variation in performance means that more relief crews are needed to maintain operations. Variation in freight service also affects time sensitive goods, connections at terminals, customer satisfaction, and equipment utilization.



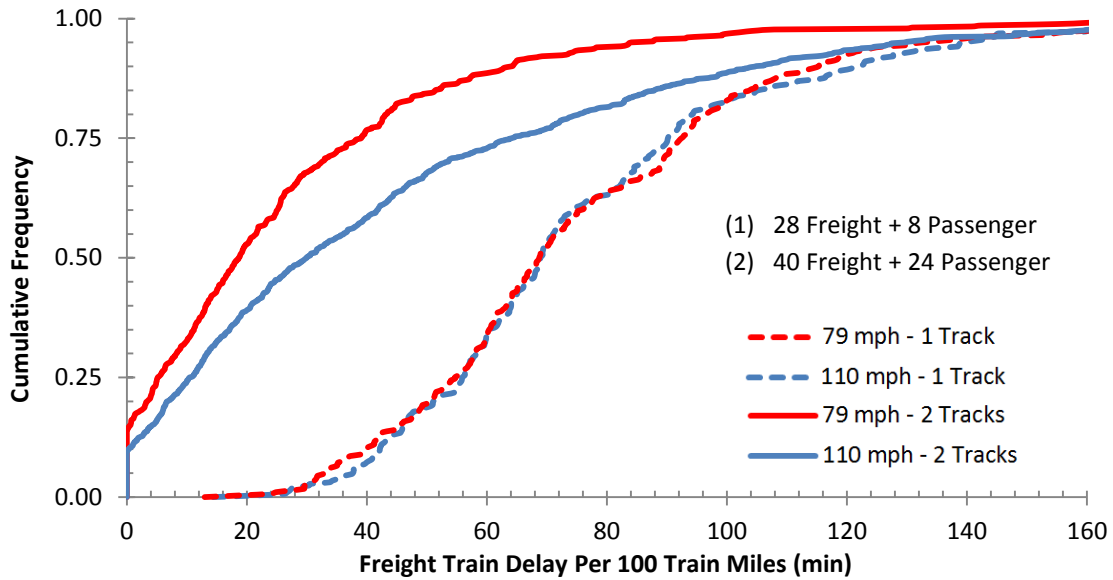
**Figure 5-4: Distribution of freight delays when (a) 79 mph and (b) 110 mph passenger trains are added to a base of 24 freight trains in a single track configuration. Distribution of freight train delays in a double track configuration when (c) 79 mph and (d) 110 mph passenger trains are added to a base of 48 freight trains per day.**

On single track, there are two counteracting factors that might explain why 79- and 110-mph passenger trains conflict with freight trains in a similar manner. The negative factor is that a faster passenger train will be more likely to cause a passing conflict with a lower priority freight train. The positive factor is that the faster passenger train will be on the freight corridor for a shorter time. Freight trains will experience less stop delay waiting in sidings for faster passenger trains to go by.



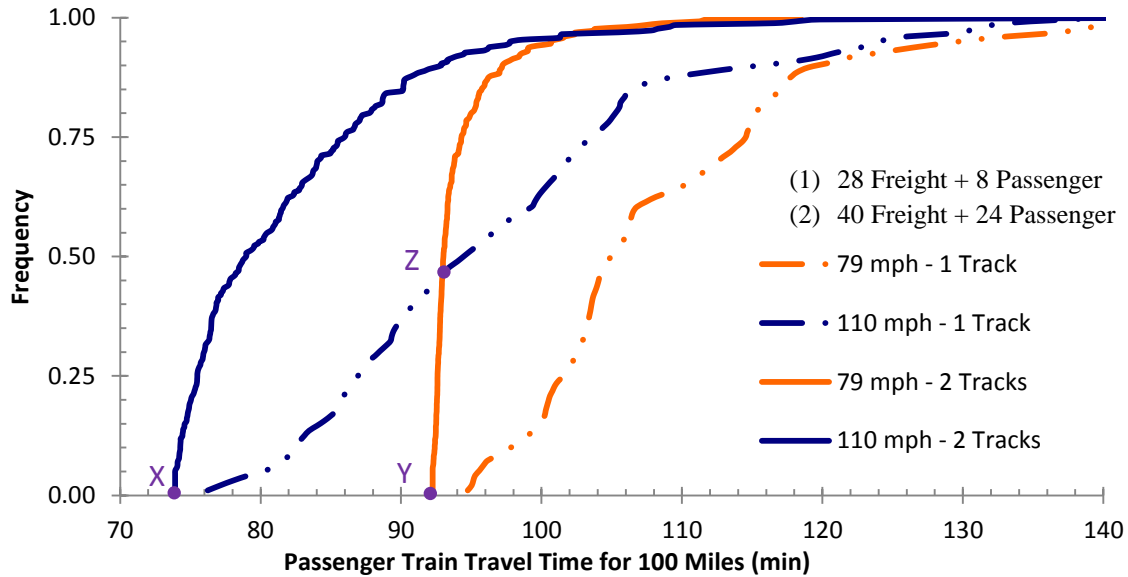
In a double track configuration, the addition of passenger trains causes the median freight train delay to increase. Figure 5-4c and Figure 5-4d show the distributions of freight delay in 10% bands. The more passenger trains operated, the greater the variation in the delay of the freight trains, and the more skewed the distribution. The 110 mph passenger trains added more delay and variability to the arrival times of the freight trains (Figure 5-4d) than the 79-mph passenger trains (Figure 5-4c). The major increase in delay starts at a lower traffic level for 110-mph passenger trains than for 79-mph passenger trains.

The cumulative frequency distributions of freight train delays between two saturated single and double track configurations are shown in Figure 5-5. The double track distributions are similar to the shape of an exponential distribution. The single-track shape is similar to a log-normal distribution with data skewed to the right. On double track, 15% of the freight trains will maintain the minimum run time with 79-mph passenger train interference. With 110-mph passenger train interference, only 7% of the trains maintain the minimum run time. The median of the double track 110 mph distribution increases from 18.7 to 30.0 minutes per 100 train miles as passenger train speed increases from 79 to 110 mph. The single track distributions for both 79- and 110-mph trains are similar, and freight trains did not maintain the minimum run time in either case. The median delays for both single-track distributions are similar.



**Figure 5-5: Cumulative frequency of freight train delays on single and double track when interfered with 79 and 110 mph passenger trains on a single track network (1) featuring 28 freight + 8 passenger trains per day and a double track network (2) featuring 40 freight + 24 passenger trains per day.**

As expected, passenger train run times per 100 miles are generally faster with 110-mph trains than with 79-mph trains (Figure 5-6). The minimum run times (MRT) are denoted by the 0th percentile and are 73.9 minutes (Y) at 110 mph, and 92.2 minutes (X) at 79 mph train speeds on double track. A steep slope from the MRT point indicates better reliability. If the data within the 5th percentile and 95th percentile are considered likely to occur, then 90% of the times to travel 100 miles are within 62.9 minutes for 110-mph track speed on double track. At a 79-mph track speed, 90% of the data are within 22.3 minutes. The 79-mph passenger train operates more consistently at the MRT than the 110-mph passenger trains. While 110 mph can offer faster travel times, the train suffers more time loss in delay events causing lower reliability (10). On single track, none of the trains perform at the MRT. Delays at sidings reduce the performance of the passenger trains. Over 50% of 79 mph passenger trains on double track have faster run times than the 110-mph passenger trains in the single-track configurations (Z).



**Figure 5-6: Cumulative frequency distribution of travel times of 79 and 110 mph passenger trains on a single track network (1) featuring 28 freight + 8 passenger trains per day and a double track network (2) featuring 40 freight + passenger 24 trains per day.**

#### 5.4 Correlation Analysis

A Spearman rank correlation test was performed on the data from both saturated single and double track networks. This measurement differs from a standard Pearson correlation coefficient where a value of +1 indicates a monotonic relationship between train performance and train speed. More data were generated to complement the previous analysis in order to increase confidence in the relationship between train performance and passenger train speed. Thirteen speeds between 50 and 110 mph were tested at 36 TPD for single track and 64 TPD for double track. Seven different traffic mixtures are considered.

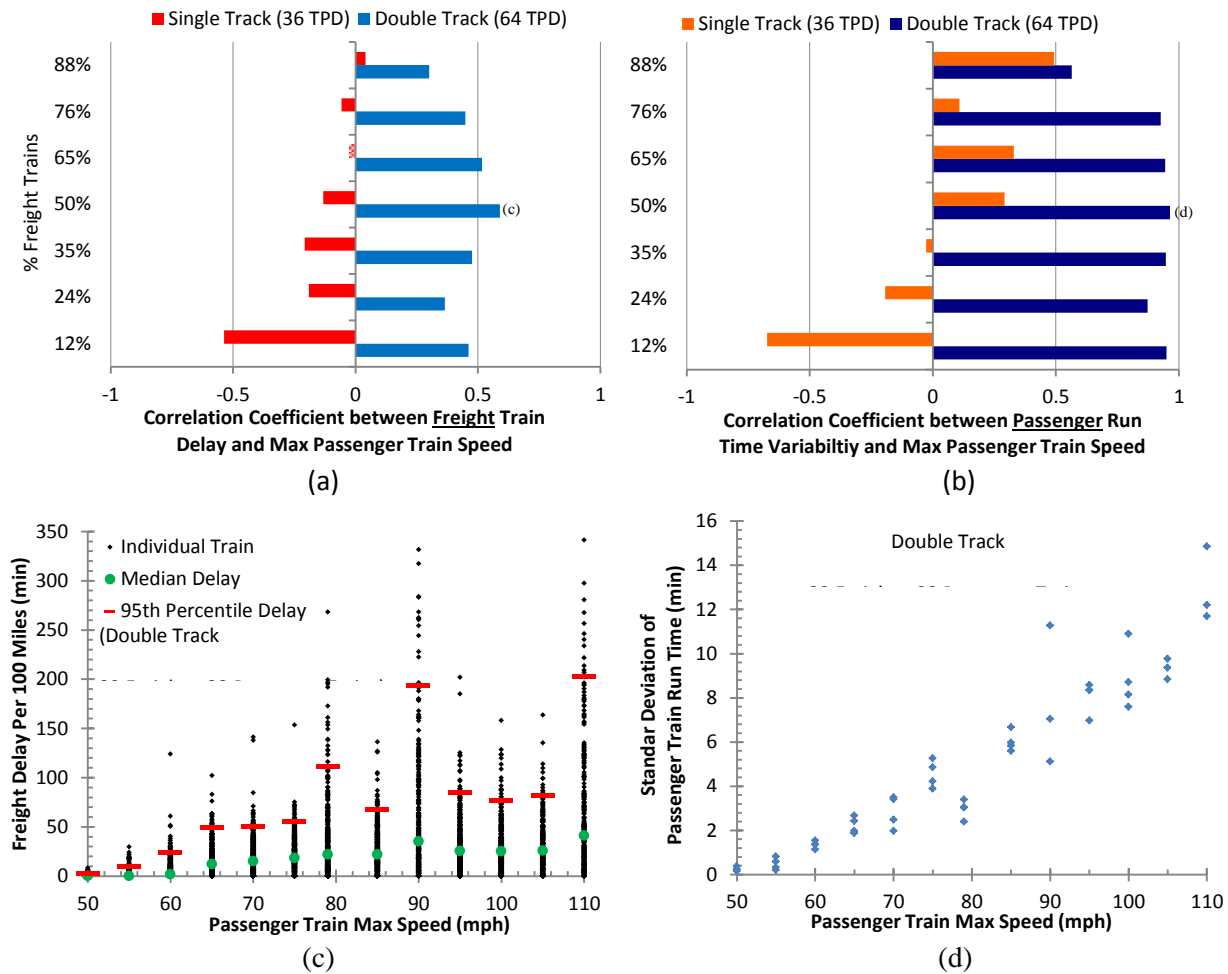


Figure 5-7: (a) Spearman correlations between maximum passenger train speed and freight train delay. (b) Spearman rank correlation between maximum passenger train speed and passenger train standard deviation of run time. Each bar in (a) and (b) corresponds to a scatterplot; examples are shown for freight train delay (c) and for passenger train reliability (d) on double track with 32 freight and 32 passenger trains per day.

Figure 5-7c shows a 50/50 traffic mixture in a double track configuration. Each data point is the delay of an individual freight train plotted at the maximum speed of the conflicting passenger trains in that particular simulation. Figure 5-7a shows seven different traffic mixtures and the corresponding Spearman correlation coefficient in both single and double track scenarios. Each bar in Figure 5-7a corresponds to a scatterplot similar to Figure 5-7c. The single track coefficients did not show positive correlations between passenger train speed and freight

train delay. There is a moderate positive correlation between passenger train speed and freight train delay on double track for the seven different traffic mixtures. The freight trains in double track are more sensitive to a large speed differential than in single track.

The previous analyses used run time to summarize the performance of the passenger trains. Correlating run time with various passenger train speeds is trivial. Instead of using run time, the analysis focused on investigating a possible correlation between reliability and passenger train speed. The reliability of the passenger trains is measured by the standard deviation of run times in a single simulation replication of a traffic mixture. Figure 5-7d shows an example scatterplot for maximum passenger speed and standard deviation of passenger train run time. There are fewer data in the passenger train correlations than the freight data because the standard deviation is a sample statistic instead of an individual train statistic. Consequently, this particular statistic varies less over 4 replicates than delays of individual trains. Each bar in Figure 5-7b corresponds to a scatterplot similar to Figure 5-7d. On single track, the results are mixed when compared with double track. In single-track, passenger-dominated networks, where passenger train delay is primarily from meets with other passenger trains, the faster train speed showed a tendency to reduce the standard deviation of run times and increase reliability. Presumably this is because each faster passenger train is on the network for a shorter time (and creating fewer meets) and there are fewer freight trains to overtake, leading to fewer sources of variation in run times. In freight-dominated single-track networks, where passenger train delay is primarily from meets and passes with freight trains that occur regardless of passenger train speed, increased passenger train speeds did not show a significant correlation with variability in the passenger train arrival times. On double track, where passenger train delay is primarily from trailing behind and overtaking freight trains, increased passenger train speeds lead to greater

variability in the run times of the passenger trains across the seven traffic mixtures. With a greater speed differential, faster passenger trains are more likely to catch up to one or more freight trains, and will experience a greater time penalty for each instance of trailing and overtaking, leading to additional sources of variance of greater magnitude compared to slower passenger trains.

## **5.5 Conclusion**

Sharing tracks with higher speed intercity passenger trains and freight trains is a challenge. In all studied cases, the addition of passenger trains to a corridor increased delay to freight trains but the mechanisms differ between single and double track configurations, freight and passenger dominated lines and passenger train speed. On single track, the greatest impact to freight trains is due to the higher priority of the passenger train, rather than its speed. The passenger train has difficulty maintaining its minimum run time in a saturated network as it is delayed primarily for meets with other passenger trains and overtaking freight trains. In a double track configuration, the meet delay is mostly eliminated and subsequently a higher speed differential between train types causes more overtake conflicts. Increasing the passenger train speed reduces the travel time, but may also decrease reliability.

In planning shared corridor infrastructure and operations, the differing characteristics of single and double track should be carefully considered. On single track, an incremental upgrade of passenger train speed from 79 mph to 110 mph may have only a modest effect on freight train delay. However, to take full advantage of the decreased run time afforded by higher speeds, planners should investigate schedules and operational strategies such as fleeting that may reduce the number of meets between passenger trains and freight trains to be overtaken. Alternatively,

the addition of double track segments to create hybrid track configurations may also mitigate the variability in run times created by these meets and overtakes. On double track, an incremental upgrade from 79 mph to 110 mph may increase freight train delay and also may increase the variability in passenger train run time. Operational strategies and track configurations, such as optimized crossover spacing, that minimize or reduce the impact of overtakes should be investigated to counteract these trends when developing operational plans and infrastructure designs for shared trackage.

## CHAPTER 6

### ANALYZING THE TRANSITION FROM DOUBLE TO SINGLE TRACK WITH DESIGN OF EXPERIMENTS TECHNIQUES

#### 6.1 Introduction

Much of the railway traffic in the United States operates on routes that consist of a single track with passing sidings. Only 37% of mainlines with annual traffic levels of ten million gross tons or more have multiple mainline tracks (Richards & Cobb 2006). With demand for freight and passenger rail service increasing, many of these single-track lines may be near capacity and will need to be upgraded. Some interim upgrade solutions to increase capacity include extending sidings to accommodate longer trains as well as adding additional sidings near the midpoints of long sections of single track. After these short-term solutions are implemented, additional sidings will offer diminishing returns and double tracking will be necessary to handle additional traffic. This second main track can be phased in over time with the amount installed matched to expected increases in rail traffic. During these intermediate phases, the track arrangements may have operational characteristics of both single and double track.

In this chapter, I describe an approach to quantify the transition from single-track-like characteristics to double-track-like characteristics. This will involve analysis of infrastructure upgrades to a high quality, single-track line with dense siding spacing, high entry speeds, and capacity for long trains. I will describe a series of simulation analyses in which sections of double track are systematically added to this base case in order to quantify the transition from single track to full two-main track. Additionally, various operating parameters are considered that may affect the capacity benefit from partial double tracking.



Many factors affect how trains perform over a railway network. The purpose of this study was to focus on a subset of these factors to determine fundamental relationships that can improve the understanding of railway performance. It is intended to illustrate the relative effects under various conditions rather than absolute measurements for a particular set of conditions. . To investigate the properties of hybrid track configurations, design of experiments (DOE) software was used to create a subset of hybrid track scenarios. The scenarios were simulated with the Rail Traffic Controller (RTC) software and multivariate regression techniques were used to quantify the main factor and interaction effects influencing train delay and line capacity.

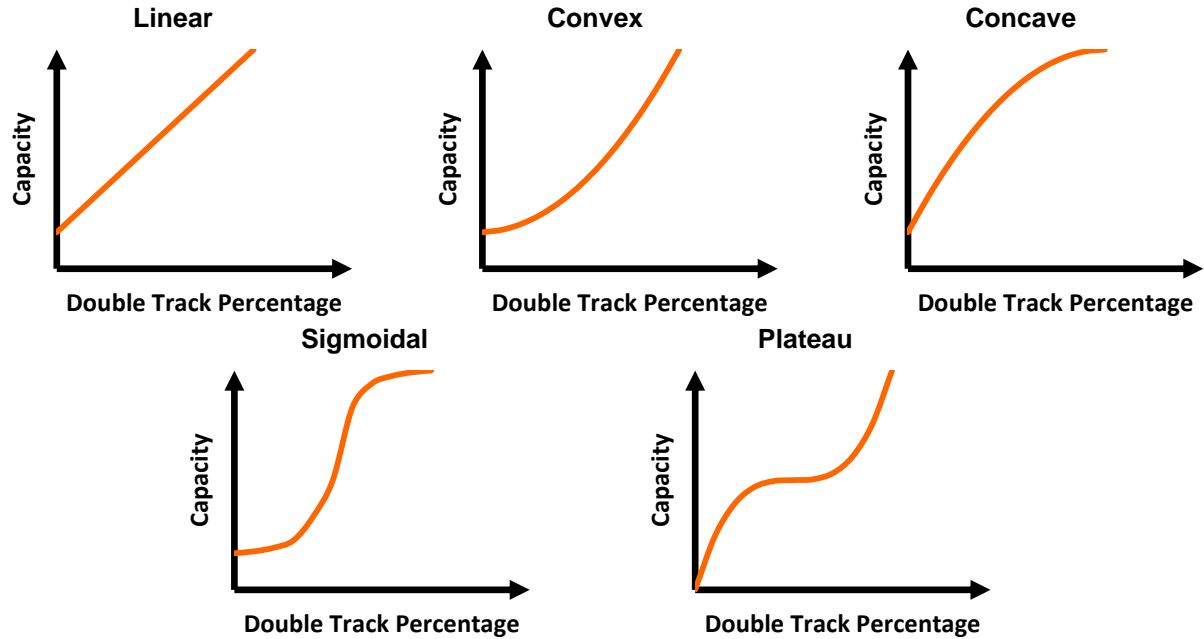
### **6.1.1 Background**

Previous research has shown that single and double-track railway lines behave differently. Single-track lines operate at have substantially less capacity than double-track lines. The primary reason for this reduction in capacity is because trains traveling in opposite directions having to alternate use of single-track sections between pairs of passing sidings. Often, the pair of adjacent sidings with the longest transit times is identified as the bottleneck (Pachl 2004). On double-track lines, the theoretical capacity is related to the following distance between trains moving in the same direction. The theoretical capacity of double track is reduced if there are speed differentials, overtakes, or traffic traveling against the current of traffic on the track normally used to move trains in the opposite direction (Kahn 1979, Sogin et al. 2013a).

Previous studies on single track determined that adding a high priority train to a freight network will increase average train delays more than simply adding another freight train of equal priority (Sogin et al. 2011). The speed differentials between trains showed very small correlations between train delay and maximum train speed. Meets at sidings are often cited as a primary delay mechanism on normal, single-track operations (Dingler et al. 2010). Double track

delay is much more sensitive to speed differentials because a faster, high-priority train may need to either run below its maximum speed, accruing delay when following a slower train, or use the second track in the opposing direction to overtake a slower train, potentially delaying oncoming trains. Because single and double track configurations have different delay response models, I wanted to understand the performance transition function between a single-track line and a double-track line as track is installed, measured by the percentage of second main track. A related question is whether the arrangement of the second track along a line affects its capacity or the functional nature of the transition from single to double track. In this chapter, double track refers to two tracks that are signaled in both directions to maximize flexibility. In some railroad operating rules this is referred to as 2MT (Pachl 2004).

A positive relationship between double-track percentage and capacity is assumed. Five hypothetical capacity-transition functions are shown in Figure 6-1. Each of these functions has an important railroad implication. A linear transition would assume that each mile of double track installed would have equal improvement to the capacity of the line. Additionally, A convex relationship would indicate that there are compounding benefits to having more second main track installed rather than less. In this case, there are greater returns on investing in second main track sections when most of the line has double track. A working hypothesis is that longer sections of second main track allows for more time to accelerate to maximum authorized speed between meet conflicts with other trains. However, a concave relationship would indicate that initial sections of second main track have the greatest returns on capacity improvement. Sigmoidal and plateau functions would indicate that there would be critical thresholds of second main track needed in order to change line capacity. Regression techniques were used to identify the capacity benefits by installing sections of double track.



**Figure 6-1: Possible double track transition functions.**

There has been limited study of hybrid track configurations. Preston and Taylor (1987) determined that the combination of careful scheduling and long sidings could be an alternative to full, two-main track. Harrod used mixed-integer optimization techniques to find low-delay dispatching solutions on hypothetical single-track corridors featuring large speed differentials and extra-long sidings. Additionally, Pawar (2011) used analytical models to determine the length of sidings required to run a single-track, high-speed railway without delays in meets. Lindfeldt (2010) compared scenarios featuring partial double track and scenarios with additional sidings and determined that partial-double track offers more flexibility to timetables and greater ability to prevent secondary delays.

Regression modeling of train delays has been used in the past to quantify effects of various operational factors on train delay. Prokopy and Rubin (1975) used single-track simulation results to develop a multivariate regression model. Krueger (1999) employed a

similar approach using an updated simulation model and also summarized the data through multivariate regression. Both models were developed by varying only one parameter at a time. Mitra et al. (2010) developed a regression model for single track with eight variables derived from simulation results. This model did not consider interaction effects between variables. Lai and Huang (2012b) used regression and neural networks to model RTC simulation results from both a single and double track network. For both the single and double-track models, they used a full-factorial experimental design analyzing five factors at three different levels.

Train delay is often estimated as an exponential function of traffic volume (Equation 6-1a) (Prokopy & Rubin 1975, Krueger 1999). This particular equation is useful because it has a closed form solution for relating volume to a level of service standard (Equation 6-1b). However, in this analysis, train delay was modeled using a multi-variable, polynomial response surface model with interaction terms. The equation defining this delay response surface does not guarantee a closed form solution for rail capacity in terms of traffic volume, but can give insight into how various factors might change the  $A$  and  $k$  parameters of Equation 6-1a.

$$D = Ae^{kV} \quad (6-1a)$$

$$V = \frac{1}{k} \cdot \ln\left(\frac{D_{max}}{A}\right) \quad (6-2b)$$

Where:

$D$  = Delay

$D_{max}$  = Average delay corresponding to the minimum level of service

$A$  = Delay mitigation constant

$k$  = Congestion factor

### 6.1.2 Factor Selection

A common capacity measurement for North American freight corridors is maximum trains per day (TPD). This value is typically tied to a minimum acceptable level of service (MLOS) requirement where additional traffic degrades the transportation utility provided to the users of

the rail service. In the analyses described in this chapter, the MLOS was measured as average train delay, which is defined to be the difference between minimum run-time (MRT) and the actual run time of a particular train type traveling over a 245-mile corridor. MRT equates to the fastest a train can traverse a route without any interference from other trains or other external factors. Large delays corresponded to a low quality level of service, and small delays corresponded to a high level of service. Once train delay is calculated for various traffic levels, capacity can be interpolated from the traffic volume corresponding to the maximum average delay ( $D_{max}$ ) used to establish the MLOS. Using this methodology, train delay is used as a proxy for maximum daily throughput (capacity).

Five factors that may have a large influence on train delays in a hybrid track configuration were investigated (Table 6-1). The different permutations of these five factors represent different operating and infrastructure conditions with a mixture of freight and passenger traffic. Larger ranges of these factor levels yield better parameter estimates. However, smaller factor ranges yield higher resolution over a smaller region of the true response surface. If the response surface is extrapolated beyond the ranges of the inputs, then it will not adequately represent the response outside this range (Kutner et al. 2004). Additionally, several possible combinations of maxima and minima of these factor ranges represent extreme operating conditions that may not be practical in actual railroad operations. However, simulating these extremes can reduce error in parameter estimates. Consequently, selection of appropriate factor ranges was critical to obtaining a response model that properly captures overall effects without being overly influenced by extreme local or boundary conditions (Schmidt & Launsby 1994, Goos & Jones 2011).

**Table 6-1: Factors studied in capacity analysis.**

Factor	Range	No. of Levels
Total Trains Per Day	0-60	5
Traffic Mixture	17% - 83%	3
Percent Double Track	19% - 100%	9
Passenger Speed (mph)	40 - 100	4
Double Track Allocation	Alternate → Split → Group	3
Responding Variable: Train Delay		

The levels for traffic volume, measured by total trains per day (TPD), are indicative of free-flow and congested networks. “Traffic mixture” reflects the traffic composition of the line, expressed as the percentage of trains on the network that were freight trains. A value of 17% describes a Passenger Dominated Corridor (PDC) where 83% of the total TPD were passenger and 13% of the traffic was freight. A mixture of 83% was a Freight Dominated Corridor (FDC). A value of 50% was evenly split between the two types of traffic. The double-track percentage referred to the amount of the route that had two tracks. This calculation included both the length of passing sidings and the length of double track. A value of 19% double track corresponded to the single-track base case and a value of 100% corresponded to a full two-main-track configuration. Without cab signals and signal enforcement systems such as automatic train stop, automatic train control or positive train control, the maximum speed of a passenger train is limited to 79-mph (Federal Railroad Administration 2013c). On developing passenger corridors across the United States, potential maximum passenger train speed upgrades are typically to 90-mph and 110-mph. Also, a minimum 40-mph passenger train speed is introduced to investigate the effect of eliminating the speed differential between train types but maintaining priority.

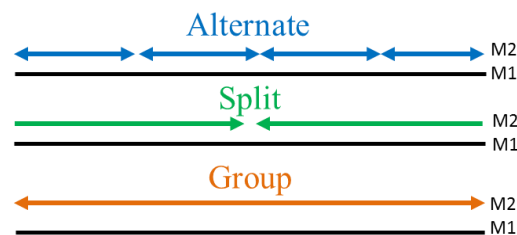
In this analysis, there were always two types of freight trains. The first was a bulk train with a maximum speed of 40-mph. The second train was an intermodal train with a maximum

speed of 60-mph. These intermodal trains were included because most FDCs still have some degree of heterogeneity between freight trains. In all simulations, there were three bulk trains for every one intermodal train. The intermodal trains were constrained to not exceed the maximum speed of the passenger trains. For permutations featuring 40-mph passenger trains, the intermodal trains were also limited to 40-mph to eliminate all speed differentials. Adding the maximum speed of the bulk and intermodal trains to the factor list of Table 6-1 would have greatly expanded the experimental design matrix and clouded the effects of the five factors of interest.

In addition to the varying operating conditions described above, from an infrastructure standpoint, there are various strategies for how the second mainline track could be constructed in phases and distributed across a single-track corridor. There will likely be sections of the route where the second track costs more to construct than on other sections. Based on a strategy of minimizing capital investment per mile of double track, the inexpensive sections of double track would be constructed before the more expensive sections. However if the costs of all double-track segments were relatively equal, or if cost was not the primary factor in the decision making process, then there might be an optimal distribution of double track from an operating perspective. It could also be that differences in capacity utility between segments may help offset the cost of more expensive segments and provide a rate of return that justifies investment in certain more expensive segments first. To investigate these possibilities, the double track allocation strategy was introduced into the experiment design as a factor variable.

Three potential double track allocation strategies are illustrated in Figure 6-2. *Group* describes the case where additional double track is consolidated in the center of the subdivision. A rationale for this approach is to provide priority trains enough second mainline track grouped

in one continuous segment to execute an overtake maneuver and regain maximum speed after slowing down for a turnout. *Alternate* is the strategy of picking four to six points on the subdivision and building out in both directions from these midpoints to create alternating sections of double and single track. Although this results in shorter double-track segments, the *Alternate* strategy increases the number of locations where “rolling-meets” can take place between two opposing trains without either one having to come to a stop. The final hybrid strategy, *Split*, is where the double-track resource is divided between the two terminals on the subdivision and built towards the middle. This strategy has the benefit of addressing potential bottleneck constraints at terminals by double-tracking these segments first. In order of increasing continuous double-track units, the ordinal set is *Alternate*, *Split*, and *Group*.



**Figure 6-2: Different strategies on how to allocate sections of second-mainline track to a single track line.**

The configurations followed a set of infrastructure standards summarized in Table 6-2. At higher double track levels, as existing sidings become incorporated into long double track segments, the control point at one end of each siding is converted to a universal crossover while the other control point is removed. This results in crossovers that are spaced at roughly 10-mile intervals on the double track segments. At 100% double track, the corresponding grouping strategies yield approximately the same route with minor variations in the exact locations of these crossovers.



**Table 6-2: Route parameters guidelines.**

Parameter	Value
Length	240 miles
Siding Spacing	10 miles
Siding length	10,000 feet
Diverging turnout speed	45 mph
Traffic control system	2-Block, 3-Aspect CTC
Average signal spacing	2.0 miles

There are 1,620 different permutations of the five factors identified in Table 6-1. Simulating all scenarios requires extensive model building and computing time. Design of Experiments (DOE) can reduce the number of scenarios to simulate by intelligently selecting a subset of the full factorial experiment design (Goos & Jones 2011). Using a regression analysis, estimates of parameters can be made with tight confidence intervals over a smaller dataset (Schmidt & Launsby 1994, Goos & Jones 2011). JMP is a commercial statistical analysis software package that is used to create partial factorial designs (JMP 2013). Additionally, JMP can create a constrained experimental design matrix. In this case, the experiment design matrix was constrained to avoid permutations with combinations of high traffic levels and low double-track percentages that are infeasible for RTC to solve and do not represent realistic routine railway operations. In order to create a response surface model for the five factors of interest, JMP suggested a design matrix with 89 different scenarios, approximately 5.5% of the full-factorial experiment.

### **6.1.3 Simulation Methodology**

Rail Traffic Controller (RTC) simulated the scenarios suggested by JMP. RTC is the de facto standard for railway simulation analyses in the United States. Users include all Class I Railroads, Amtrak, BART, and many others. An RTC simulation run for a particular scenario simulated

five days of traffic on the corridor. Each simulation run was then repeated eight times to yield performance statistics for 40 days of railway traffic for each of the 89 scenarios in the experimental matrix. Replication of each simulation run has several benefits. Most importantly, if a particular randomized simulation run was infeasible for RTC to dispatch, it could be dropped from the analysis with the remaining feasible replicates for that scenario still available for use in the final analysis. Additionally, replication and randomization creates the opportunity for the RTC simulation algorithm to make different decisions under a similar set of inputs.

In addition to the physical and operating characteristics identified in Table 6-1, the schedule of traffic affects corridor performance. In many countries, conflicts between trains are carefully planned and managed with a timetable (Hallowell & Harker 1998). However, railway operations in North America are highly variable due to fluctuations in freight traffic demands, weather, and other sources of variation and delay. Dispatchers are resolving conflicts between trains in real time instead of following a strict timetable (Ireland et al. 2003). Different schedules for the same number of trains on the same infrastructure may yield considerably different average train delay metrics. Consequently, assuming a certain schedule for any given scenario may result in large experimental error, which will lead to biased results. In order to reduce error due to schedule bias, the daily departure time of each train was determined using a random uniform distribution over a 24-hour period. This process created a new schedule for each simulated day. “Good” low-delay schedules were averaged against “bad” high-delay schedules over the 40-days of simulated traffic. Stable averages can be achieved by averaging train performance over multiple days.

Each train type in the simulation is based on the characteristics specified in Table 6-3. The bulk and intermodal train characteristics are based on the Cambridge Systematics National

Rail Freight Infrastructure Capacity and Investment Study (2007) conducted for the Association of American Railroads. Freight car gross rail load and length are based on averages for each car type. The horsepower-per-trailing-ton ratios are based on experience and information from the 2002 Transportation Research Board Workshop on Railroad Capacity and Corridor Planning (Galloway 2002). The passenger train is based on the 110-mph operation between Chicago and Detroit. The acceleration curves from the simulation model were matched to GPS coordinate data from the Amtrak geometry car on this corridor. The passenger train stops were spaced at 32.4-mile intervals based on the current Amtrak station spacing on routes in California, Illinois, Washington, and Wisconsin.

**Table 6-3: Train Parameters for Simulation Model.**

	Bulk	Intermodal	Passenger Train
Locomotives	3 SD70 locomotives	5 SD70 locomotives	2 P42
No. of Cars	115 hopper cars	75 cars	8 Single level cars
Length (ft.)	6,325	5,659	500
Weight (tons)	16,445	5,900	800
HP/TT	0.78	3.64	15.4
			32.4 miles between stops

## 6.2 Freight Model

After the 89 scenarios were simulated, a series of multivariate regression models were developed using the results. These models take the input parameters from Table 6-1 and construct a polynomial function that predicts average bulk-train delay. From this polynomial model, capacity was calculated by determining the traffic volume level that corresponded to the allowable average bulk-train delay specified by the MLOS. Overall, there were 34,752 data points from individual bulk trains available for the regression analysis.

### 6.2.1 Model Creation

There are four classes of variables considered for building a multivariate regression model using the simulated results: main effects, curvature, continuous interactions, and nominal-interactions. Main effects are the corresponding first-order variables representing the factor levels in Table 6-1. The grouping strategy is considered an ordinal set and has two dummy-indicator variables to represent this factor. Curvature terms are second-order quadratic terms that are only used for the continuous factors in Table 6-1. Only second-order, two-way interaction parameters are considered in this analysis (except for *Double-track percentage*  $\times$  *Volume*<sup>2</sup>). An additive interaction occurs when an increase in two variables corresponds to a change in train delay greater than what the main effects and curvature terms would independently estimate. An interference interaction occurs when two variables move in opposite directions and the corresponding change in train delay is greater than what the main effects and curvature terms would indicate. A nominal interaction refers to an interaction effect between a particular grouping strategy and one of the continuous variables (Kutner et al. 2004).

Stepwise modeling techniques selected an elite group of candidate models. These models were scored against each other based on model simplicity and fit of the data. In general, a simpler model will have tighter parameter confidence intervals than a more complicated model, even though a more complicated model may fit the data better. In this dataset, most of the variation was explained by a few parameters and the candidate models differed only in the minor two-way interaction parameters picked by the selection algorithm. In this case, simpler models still had nearly as good fit to the data as the more complicated models. The maximum R-squared that could be achieved was not 1.0 due to the replication of runs in the experimental design matrix. Under these particular replication conditions, the maximum achievable R-square

was 0.5519. The final model had an R-squared of 0.5419. All parameter estimates were significant at the 95% confidence level (Table 6-4).

**Table 6-4: Freight model parameter estimates.**

Term	Estimate	Std Error	t Ratio	t Ratio	Prob> t
Volume	129.24253	1.132794	114.09		<.0001*
Double Track*Volume	-111.7918	1.065417	-104.9		<.0001*
Double Track	-80.25624	0.960646	-83.54		<.0001*
Volume*Volume	55.243603	1.148509	48.10		<.0001*
Volume*Volume*Double Track	-47.18382	1.36686	-34.52		<.0001*
Allocation[G-S]*Volume	-21.65779	1.322958	-16.37		<.0001*
Mixture (% Freight Trains)	-12.00522	0.899088	-13.35		<.0001*
Allocation[S-A]*Volume	16.372263	1.267921	12.91		<.0001*
Volume*Hetero	-8.053779	0.630294	-12.78		<.0001*
Max. Passenger Train Speed	8.5234311	0.722078	11.80		<.0001*
Allocation[S-A]*Pass Speed	-13.84208	1.182322	-11.71		<.0001*
Allocation[G-S]	-10.51816	1.05799	-9.94		<.0001*
Allocation[S-A]	8.8705107	1.00342	8.84		<.0001*
Allocation[G-S]*Double Track	12.473623	1.420156	8.78		<.0001*
Allocation[G-S]*Pass Speed	8.8017065	1.136256	7.75		<.0001*
Allocation[S-A]*Double Track	-10.68652	1.410454	-7.58		<.0001*
Mixture*Mixture	-4.658955	0.827374	-5.63		<.0001*
Allocation[G-S]*Mixture	-6.70368	1.19551	-5.61		<.0001*
Allocation[S-A]*Mixture	6.4877827	1.181209	5.49		<.0001*
Double Track*Pass Speed	2.6273965	0.589281	4.46		<.0001*
Pass Speed*Pass Speed	2.6131413	0.742248	3.52		0.0004*
Mixture*Pass Speed	-1.856641	0.719021	-2.58		0.0098*

## 6.2.2 Main Effects

The parameters in the delay model are indicators of how train delay might change for a unit increase of each independent variable. The signs of the continuous variable parameter estimates follow current understandings of railway performance. Adding more double track reduces delay,; adding more trains will increase delay. Additionally, as the percentage of freight trains increased, they performed better in a FDC than a PDC. Higher passenger train speeds have a small effect on increasing freight train delays. With this delay model, capacity can be estimated by determining the traffic volume that corresponds to the MLOS (Allowable Delay).

The ranges of the independent variables represent a broad set of operating conditions but do not necessarily facilitate direct comparisons between each other due to differing scales and units. For example, the change in capacity over a change in traffic mixture is a different rate measurement than the change in capacity per mile of double track. In order to correct for this unit-bias, arc elasticity (Equation 6-2) is used to show how the response surface reacts to normalized unit changes in a particular factor from a base case scenario. Equation 6-2 calculates a dimensionless parameter that is independent of the measured units of the Y and X variables. For this elasticity analysis, four out of the five factors were held constant while one factor was varied and the delay and capacity response observed. The elasticity for an individual factor was calculated twice. The first calculation determined the maximum capacity increase from the base case by manipulating only the factor of interest. This calculation was then repeated to determine the change in capacity when the factor changes from its base level to the level that yields the greatest capacity reduction. For example, a change from 60% double track to 100% double track changed the capacity from 27 TPD to 69 TPD. The corresponding arc elasticity was +1.7. Conversely, a reduction to 19% decreased capacity to 18.5 TPD with an elasticity of -0.57

$$e = \frac{\Delta Y}{\Delta X_i} \cdot \frac{X_{i_o} + X_{i_f}}{Y_o + Y_{i_f}} \quad (6-2)$$

Where:

$e$  = arc elasticity

$Y$  = dependent variable

$X_i$  = independent variable i

$(X_{i_o}, Y_o)$  = Base case conditions

$(X_{i_f}, Y_f)$  = Final conditions

This particular model will have non-constant elasticities due to the intercept, curvature, and interaction terms. The small changes in elasticity as a function of the independent variables are assumed to be less than the relative global differences between variables. This bias is also reduced by calculating the elasticities twice over positive and negative changes from the base case.

The maximum delay changes and elasticities for the continuous variables are shown in Figure 6-3. Capacity was most sensitive to the amount of double track. Changing the MLOS also had a great effect on rail capacity. Increasing the MLOS from 60 minutes to 120 minutes increased capacity to 39.5 TPD. Conversely, a stricter MLOS of 30 minutes decreased the practical capacity to 15 TPD. The next most responsive main effect variable was the traffic mixture of the line. Passenger train maximum speed showed modest effects on capacity. However, higher passenger train speeds reduced capacity more than a PDC.

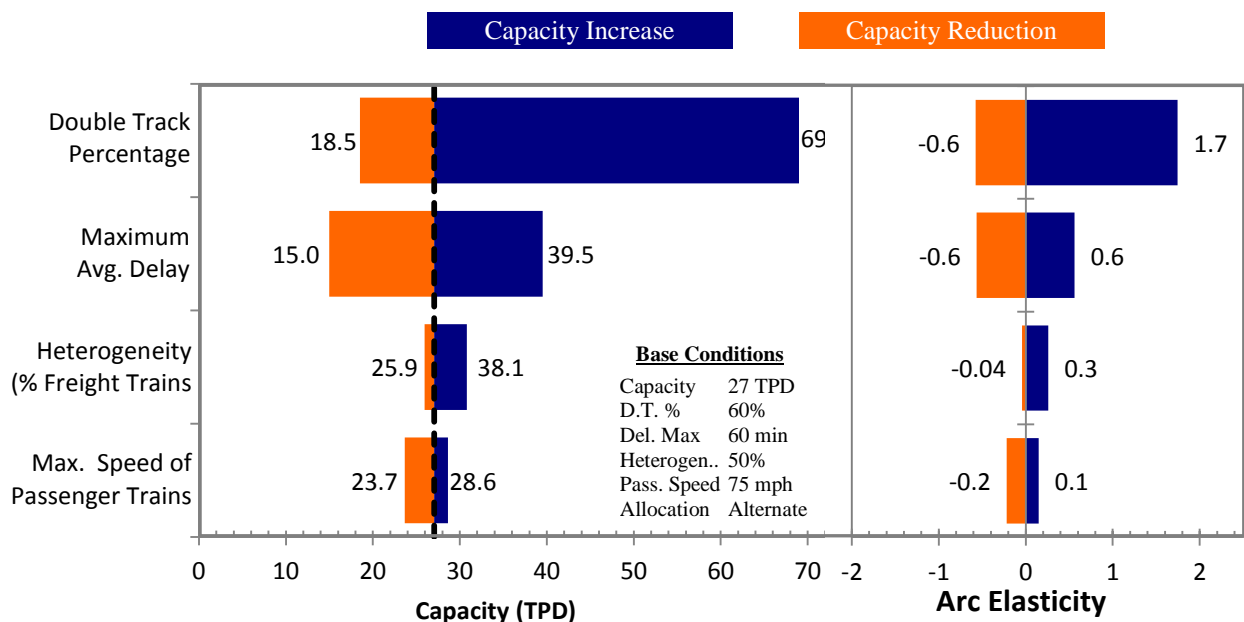


Figure 6-3: Main effects by manipulating 1 factor to achieve a maximum capacity reduction or capacity increase. Absolute changes in capacity are on the left and arc elasticities on the right.

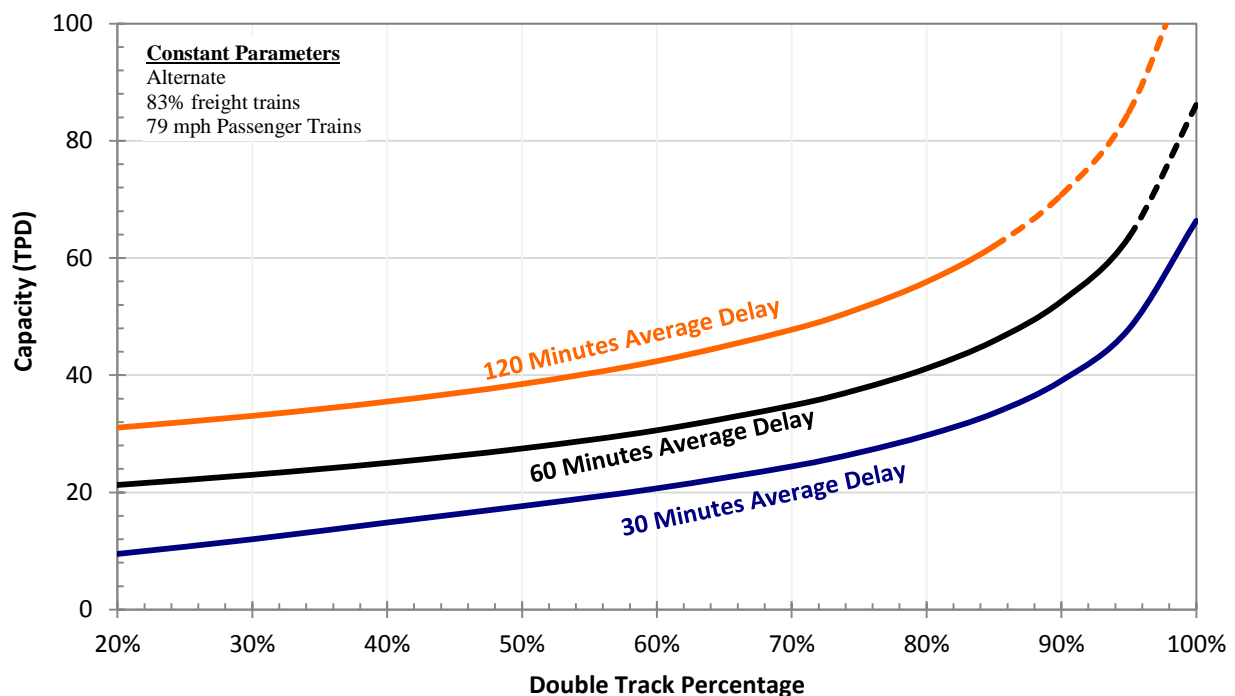
### 6.2.3 Interaction Effects

The final model included two-way interaction terms. There can be greater changes in capacity if two variables are manipulated simultaneously instead of one at a time as shown in the previous section. An important interference interaction was between traffic-volume and double-track percentage. As these two variables moved in opposite directions, there was a greater change in capacity than changing one variable independently. Figure 6-4 shows capacity as a function of double-track percentage for three different MLOS-delay standards. When the percentage of double-track was low, the capacity return on investment in track was also low and close to constant. As the percentage of double track increased, the return on capacity began to increase exponentially at about 70%. The same trend was observed for three different delay standards. The capacity contours were spaced further apart when the delay standard was small. At 60% double track, the difference between a 30-minute MLOS and a 60-minute MLOS was approximately 10 TPD over a 30-minute change (0.33 train/min). The difference between a 60-minute and a 120-minute MLOS was approximately 12 TPD over a 60-minute change (0.17 train/min). Tightening the MLOS will show a greater change in the relative vertical location of this capacity contour than relaxing the MLOS ( $0.33 > 0.17$ ). By inspection, the instantaneous slope of each capacity contour appears to be parallel (or constant) at the same double-track percentage. If this relationship holds, then the change in capacity per change in double-track percentage is independent of the MLOS.

The amount of double track required to mitigate train delay can be calculated from Figure 6-4. For example, assume that the railroad had a 50% double track corridor and wanted to change its MLOS from 120 to 60 minutes. To maintain its current operation of 38 TPD corresponding to a MLOS of 120-minutes, the railroad must add enough sections of double track to reach the 60-



minute MLOS contour which occurs at 76.4% double track. Lastly, the dashed lines of Figure 6-4 indicate extrapolation of the response surface model because none of the simulations featured traffic levels greater than 60 TPD. The interaction between double track and capacity will be the basis of the subsequent analysis discussion. In Figures 6-4 through 6-7, these two factors are shown on the axes with one additional factor of interest shown on the contours while remaining three factors are held constant.



**Figure 6-4: Change in capacity by adding sections of double track at different delay standards.**

The double track allocation strategy had a modest effect on capacity (Figure 6-5). In this model, *Alternate* and *Group* performed better than the *Split* strategy for every level of double tracking. This result may have occurred due to the *Split* strategy yielding the longest single-track section of the three approaches. If there was a complex meet between multiple groups of eastbound and westbound trains, then trains in one direction would have to be held at sidings (Kraft 1982). However, the differences in mitigation strategies may also be an artifact of the

partial factorial. After 70% double track, the *group* allocation strategy showed slightly higher capacity than the *alternate* strategy.

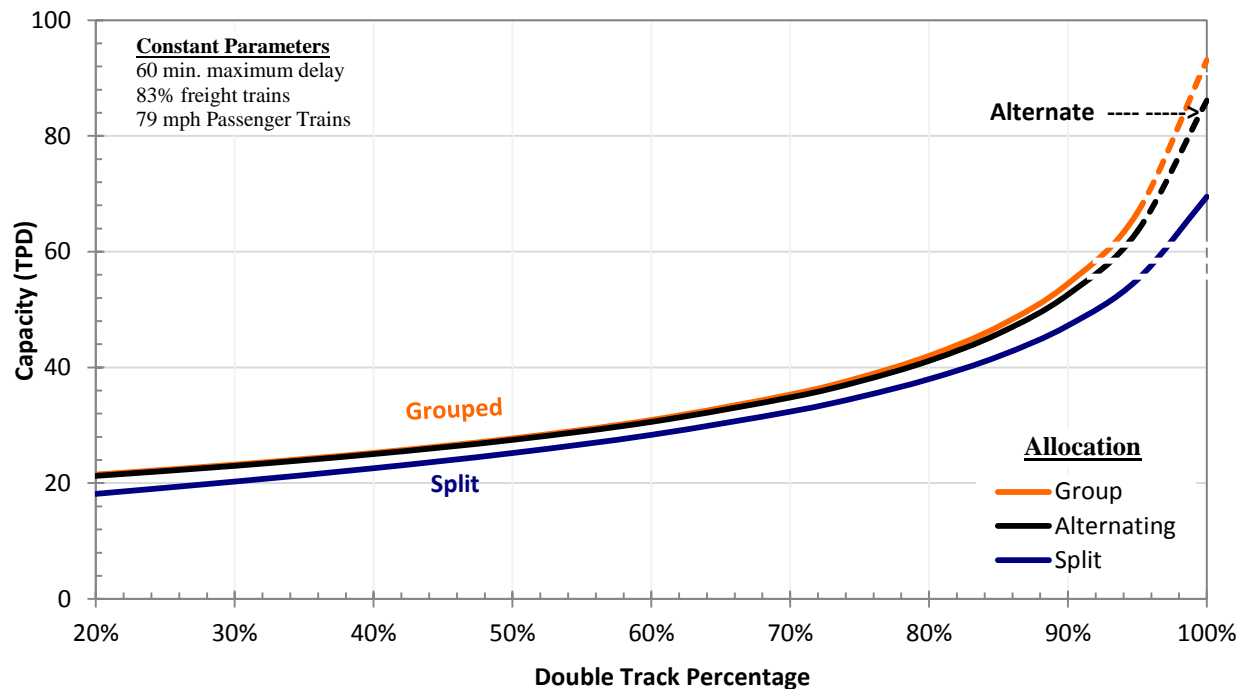


Figure 6-5: Capacity benefits by adding sections of double track by using different allocation strategies.

Heterogeneity does have a considerable impact on capacity (Figure 6-6). Freight trains perform better in FDCs (83%) than in PDC (17%). Additionally, there is a curvilinear effect in this model. The difference between corridors with 50% and 17% freight trains is small compared to the difference between 50% and 83% freight trains. In a single-track corridor (19% double track), heterogeneity changed the capacity by approximately three TPD between 17% and 83% freight trains. On a 60% double-track corridor, this range in traffic mixture corresponded to a change in five TPD. At 100% double track, this range in traffic mixture corresponded to a 28 TPD change in capacity. The impact of train-type heterogeneity increased with more double track installed.

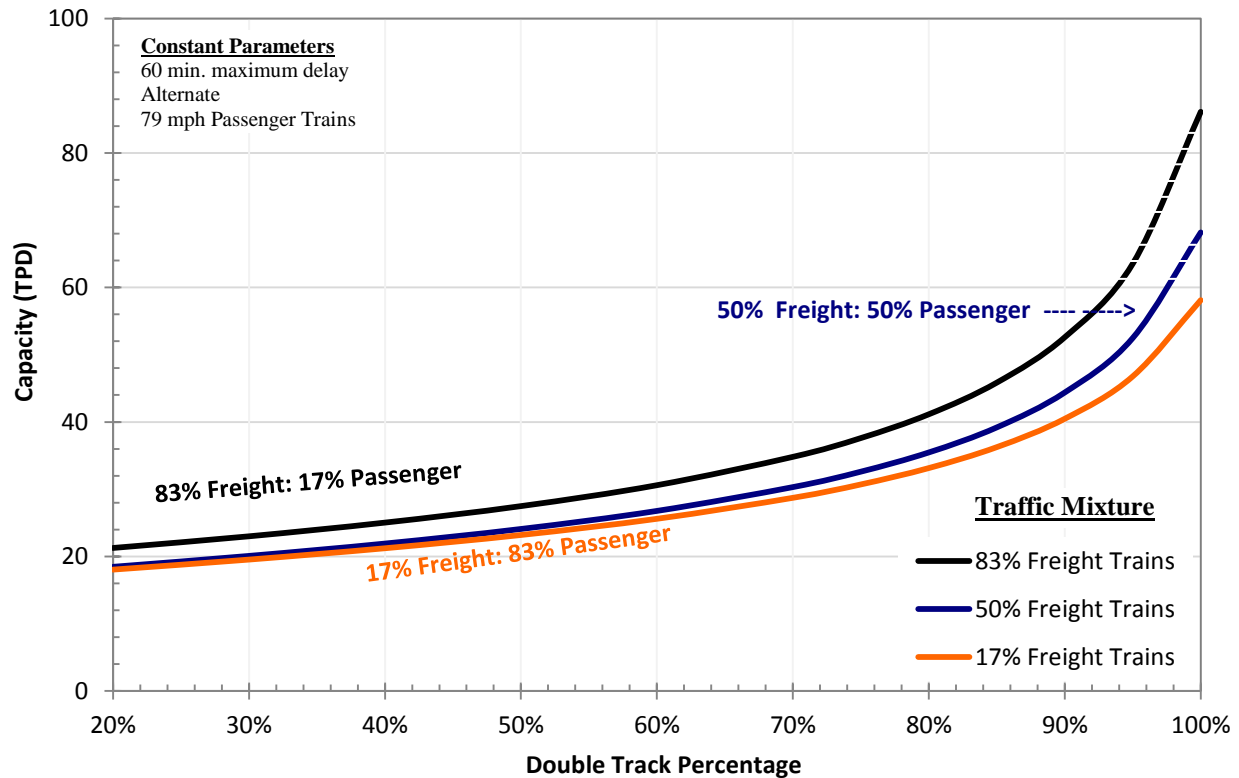


Figure 6-6: Change in capacity by adding sections of double track under different constant traffic compositions.

The maximum speed of passenger train had a modest effect on capacity (Figure 6-7). A larger speed differential in train speeds resulted in a decrease in capacity for all double tracking levels. The magnitude of this capacity reduction was greater when a greater portion of the corridor was double tracked. There was a minor interaction between maximum passenger train speed and traffic mixture (Figure 6-8: note axes changes). The 110-mph passenger trains compounded the effect of heterogeneity more than a 79-mph passenger train. Ideally, these three curves would intersect at 100% freight trains where passenger-train heterogeneity is absent as a delay mechanism. These curves have a discontinuity at 0% freight trains (i.e. a PDC) because there are no freight trains to be delayed in this traffic mixture.

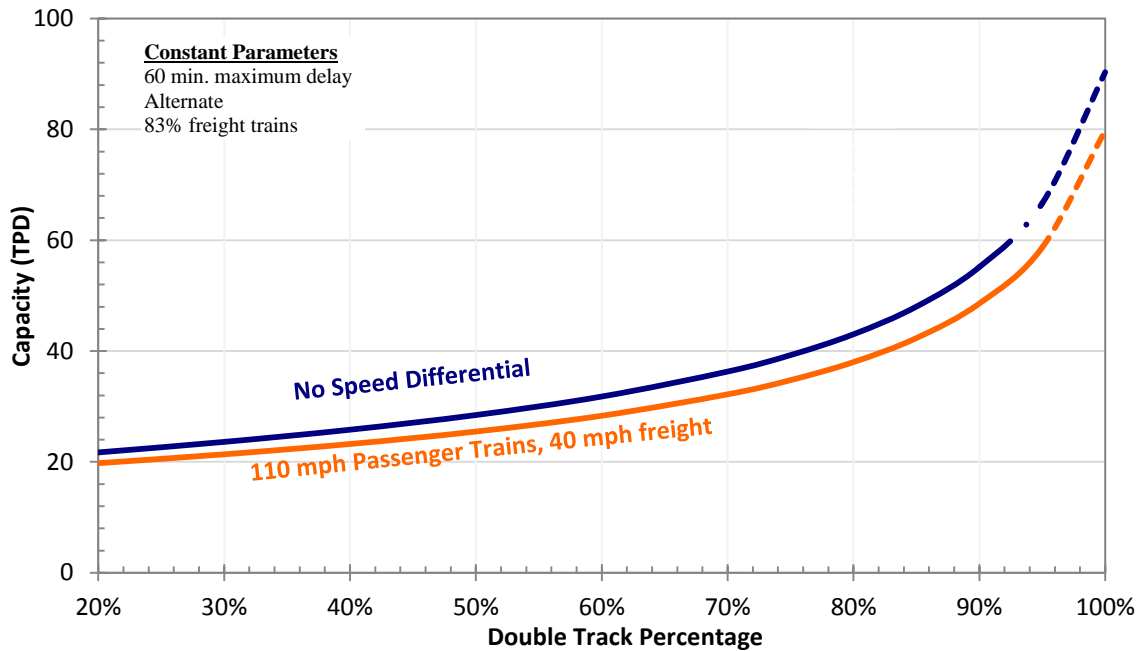


Figure 6-7: Reduction in capacity benefits by having a speed differential between train types.

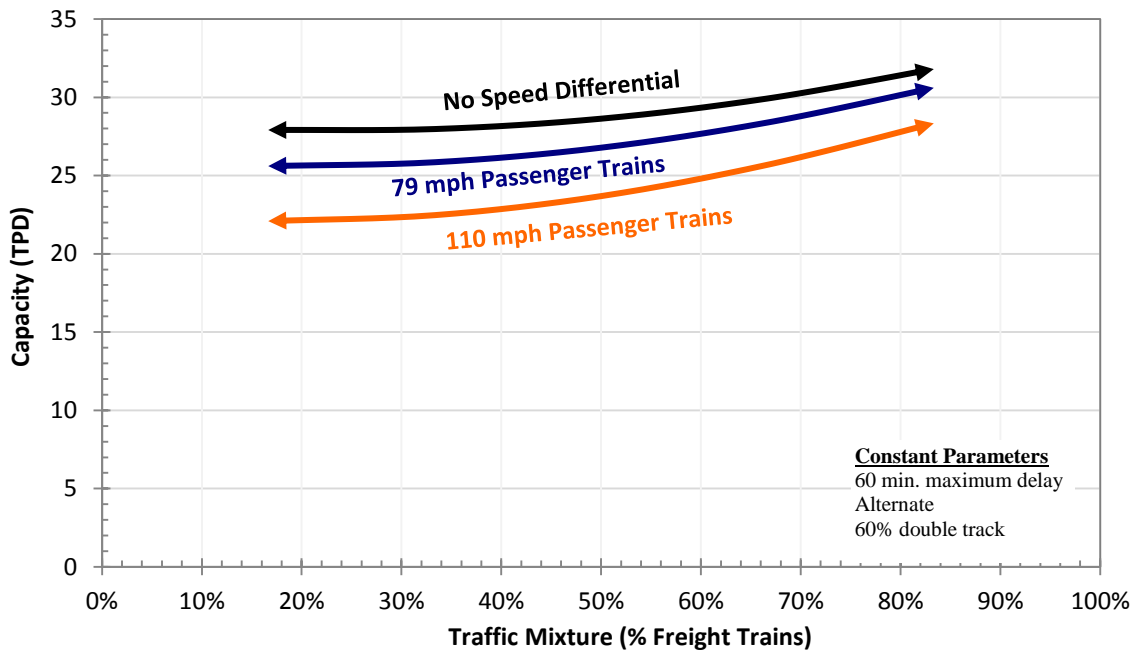


Figure 6-8: Interaction effect of maximum passenger train speed & traffic mixture on capacity.

#### 6.2.4 Model Validation

Regression models tend to fit well to the data that were used to create the model but have limited utility unless they also accurately predict “unseen” data. To validate the response surface model, 25 additional scenarios were simulated that were not selected in the original partial-factorial design. The response surface model should be able to predict these cases with the same accuracy that it predicts the original data used in fitting the model. The mean square prediction error (MSPR) should be about the same magnitude as the mean square error (MSE) of the regression model. This is an indication that the model is unbiased and can have reasonable predictive ability (Kutner et al. 2004). The MSPR for the validation set was 2,682 and the MSE of the response surface model was 3,068. Since the MSE and MSPR are the same magnitude, the model was considered valid for factor screening.

$$MSPR = \frac{\sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2}{n^*} \quad (6-3)$$

Where:

$Y_i$  = value of the response variable at the  $i^{\text{th}}$  case in the validation set

$\hat{Y}_i$  = predicted value of the response variable at the  $i^{\text{th}}$  case in the model building set

$n^*$  = size of the validation set

#### 6.3 Passenger Model

The previous analysis only focused on bulk freight train delays and capacity under a bulk freight MLOS. Passenger trains also experienced delays in the simulations. A preliminary passenger model was developed using the same methodology employed to create the freight model. This model also assumed delay as a polynomial function of the factors in Table 6-1. A disadvantage of this model is that MLOS became harder to define when maximum speeds were changing simultaneously with other factors. Delays to passenger trains traveling at different speeds have

different distance penalties. From the passengers' perspective, a 110-mph passenger train that is 15-minutes late will typically be preferable to a 15-minute late 79-mph passenger train from the passenger perspective. Lastly, an analysis of run time is insufficient because there is little overlap between the run times of 40-mph and 110-mph passenger trains.

The shape of the capacity contours (Figure 6-9) were similar to the freight contours in Figure 6-4, but had much longer linear portions at shallower slopes. The compounding benefits of double-track were smaller in the passenger model. Additionally, the passenger train models suggested higher capacities than the freight model on corridors with a low double-track percentage. Since passenger trains are the priority traffic, they were less likely to be subject to meet delay on single track. Thus it will take a higher traffic volume before passenger trains experience delays approaching the MLOS, resulting in a higher capacity compared to that defined by the freight standard for low double-track percentages. The shallow slope of these contours suggested a large change in capacity by changing the MLOS. At 60% double track, changing the MLOS from 30-minutes to 60-minutes increases the capacity by 18 TPD. Completing the double track on the corridor approximately matched this increase by changing the capacity by 19 TPD. In the freight model, the marginal benefits of completing the double track were approximately three times greater than doubling the MLOS.

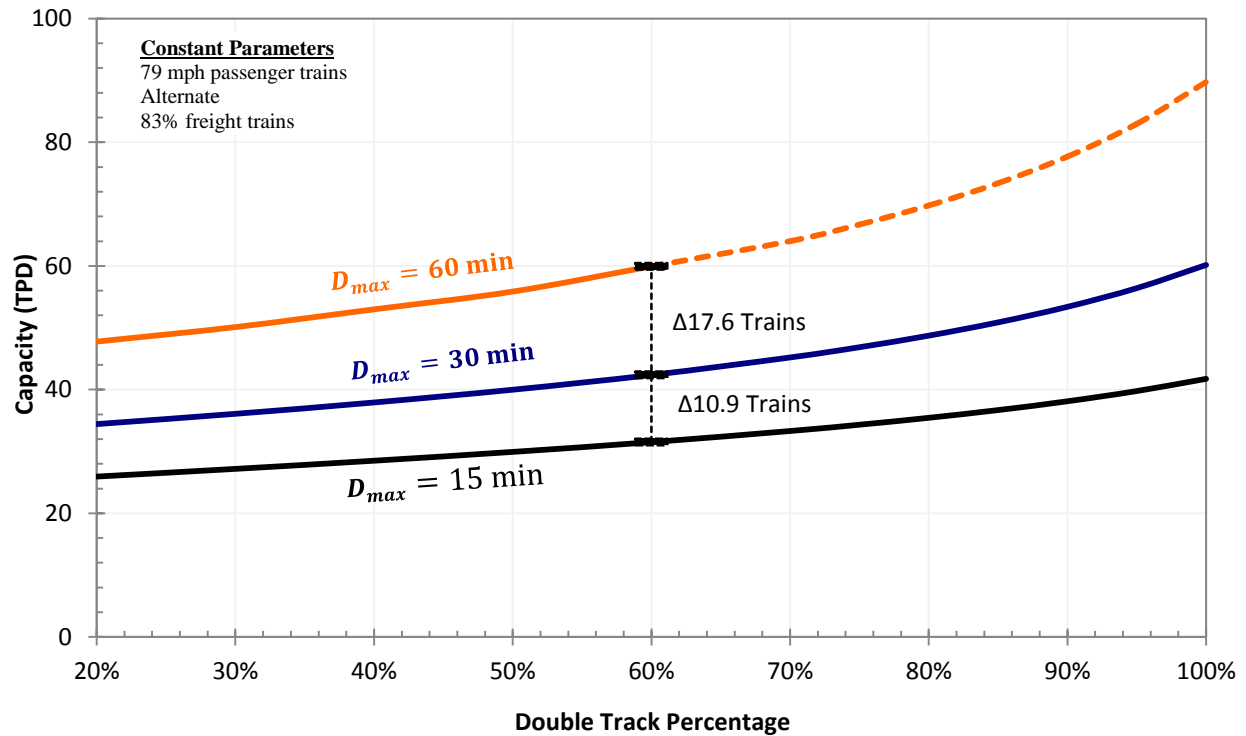


Figure 6-9: Change in capacity under different delay standards.

In a FDC, increasing the passenger train speed resulted in a significant decrease in capacity under a 15-minute MLOS (Figure 6-10). On single track, these trains had similar delays. On double-track corridors, the faster passenger trains were delayed more, and the model suggested a reduction in capacity. At 110-mph, under a 15-minute MLOS, there was minimal capacity benefit to adding double-track. Higher priority is an operational strategy to limit meet delays similar to how adding double track is an infrastructure solution to eliminating meet delays. Passenger trains already have much of their meet delay mitigated by their priority.

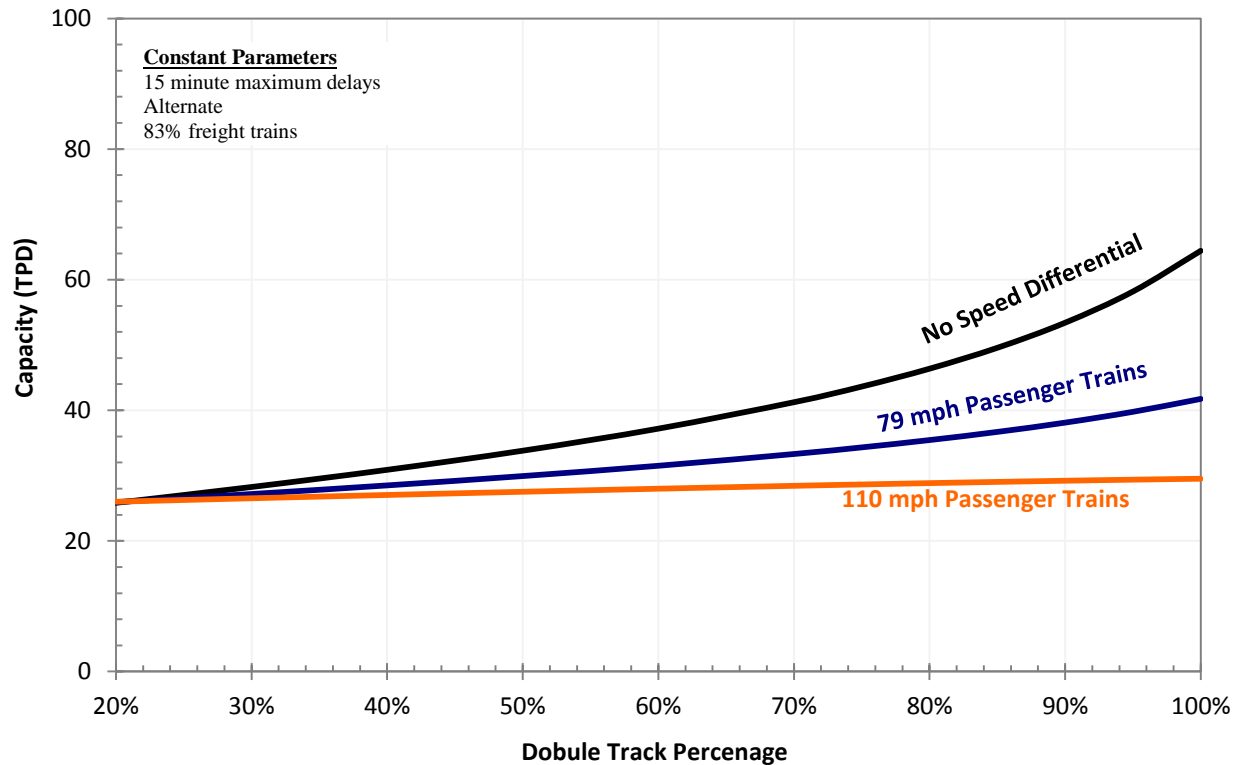


Figure 6-10: Change in capacity under different maximum passenger train speeds.

## 6.4 Combined Model

If multiple train types are considered simultaneously then the MLOS needs to be clearly defined for both train types. Once delay models are developed for each train type, then one train type model will show a lower operating capacity for a given set of input parameters than the other model. The capacity of the corridor is determined as the minimum value between the two models. Figure 6-11 shows a bulk-freight-capacity contour under a 60-minute MLOS and a passenger-capacity contour under a 15-minute MLOS. The bulk-freight-capacity contour governed the capacity of the line between 19% and 65% double track. In this range, the double-track sections were mostly mitigating freight-train delays. Between 65% and 100%, the passenger-capacity contour governed. In this range, the principal effect of the double track was



to mitigate passenger train delays. Above 90% double track, bulk-freight-train delays have dropped below 30-minutes, well under a 60-minute MLOS. With regards to return on investment, there is a non-differentiable point at 65% double track where the two capacity contours intersect. This intersection point was heavily influenced by the assumed MLOS of the two traffic types.

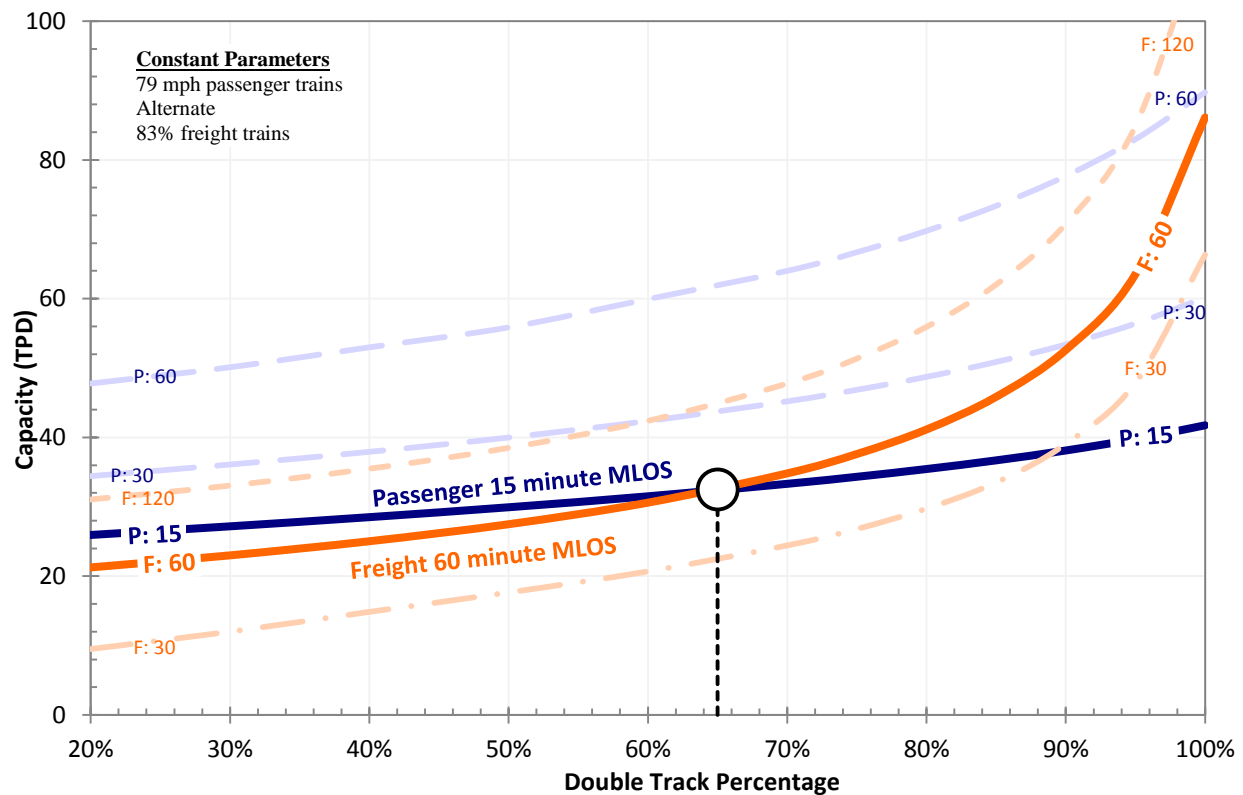


Figure 6-11: Comparing capacity contours from freight model (orange) and passenger model (blue). The dashed lines represent alternative MLOS standards.

## 6.5 Future Work

There was extrapolation in capacity beyond 60 TPD. The constrained DOE model should have allowed for more trains per day at levels of double tracking. Additionally, a 100% double-track line with homogenous traffic should have no delays under the traffic levels used for this analysis

(Sogin et al. 2013a). By only looking in the range of 17% to 83% freight trains, the effect of heterogeneity may be underestimated by not including the well performing homogenous scenarios. Also, every traffic mixture in this study incorporated some heterogeneity by always having a ratio for three bulk trains per 1 intermodal train. The differences in grouping strategy were small and may be due to the partial factorial design.

Regarding grouping strategy, one of the stated reasons for pursuing a *split* strategy is to add double track near terminals where congestion often occurs. The RTC simulation model has limited ability to account for yard congestion and the ensuing delay when trains are held outside terminals. A simulation model that specifically addresses mainline-yard interaction may reveal additional delay reduction benefits of double track near terminals. These delay reduction benefits that are not captured in the current mainline RTC model may place the *split* strategy in a more favorable position relative to the other strategies.

A model of intermodal-train delays will be developed in the future. A regression analysis of the individual intermodal-train delays could give further insight into how the three train types considered here interact. Additionally, the methodology of analyzing two types of traffic with two capacity contours could easily be applied to three different types of traffic. The resulting graph would be similar to Figure 6-11 but have two or more non-differentiable points instead of only one.

The methodology presented here could be applied to the installation of triple track. Ideally, the curvilinear relationships defined by the response surface model developed through this research would extend beyond 100% double track where values above one hundred correspond to the percent of the line where a third main track has been installed. Infrastructure

configurations to consider include parallel versus universal crossovers, overtaking sidings versus partial third-main track, and center sidings versus external sidings.

## **6.6 Conclusion**

Partial-factorial and response surface modeling can be used to quickly compare relative effects in railway simulation analysis. If a single-track corridor is being upgraded to a double-track corridor, major factors influencing the capacity included the amount of double track installed, and the MLOS. The distribution of the double track along the route can have a modest effect on train delays. Minor effects included the passenger train speed and the heterogeneity of the line. The capacity transition from single to double track was expected to be a convex (concave upward) function with greater marginal capacity benefits occurring when most of the line was double tracked. The last miles invested in double tracking can be expected to have greater returns on increasing capacity of the line.

## CHAPTER 7

### ANALYZING THE TRANSITION FROM SINGLE TO DOUBLE TRACK WITH NON-LINEAR REGRESSION ANALYSIS

*Sogin, Samuel L, Yung-Cheng Rex Lai, C Tyler Dick, and Christopher P L Barkan. (2013). "Analyzing the Transition from Single to Double Track" Submitted to the proceedings The 5th International Seminar on Railway Operations Modelling and Analysis. Copenhagen, In Press.*

#### 7.1 Introduction

There are many factors that may determine how trains perform over a railway network. Railway simulation software continues to become more sophisticated in order to better emulate actual operations. The purpose of this study was to conduct a systematic analysis using controlled experiments focused on a subset of these factors.

Most of the United States railway network is single-track with passing sidings. Only 37% of mainlines with 10 million gross tons or more are multiple-mainline-track territory (Richards & Cobb 2006). Future demand for increased freight and passenger rail service will require more capacity. Consequently, many of these single-track railway lines will need additional tracks to accommodate this demand. There are three basic approaches: adding passing sidings, extending the siding length, and adding double track. Extending passing sidings enables longer freight trains and reduces passenger delays due to meets with other trains. Additional sidings would typically be installed in the longest single-track bottleneck sections. After these intermediate solutions are implemented, double track may be the most effective way to handle the additional traffic. This second mainline track can be phased in over time such that the amount installed matches to the expected increases in rail traffic (Lindfeldt 2012b). These intermediate phases with partial double tracking will be referred to in this chapter as “hybrid” track configurations.

This analysis will focus on the capacity benefits of a single-track route transitioning to a two-main-track route in the context of shared passenger and freight train operation. Railway traffic simulation software was used to evaluate each intermediate phase at different traffic levels. A typical North American single-track route was used as a baseline condition. Sections of double track were systemically added to the base condition by connecting pairs of pre-existing passing sidings. Train delay and capacity transition curves are then mathematically described. This procedure was used first with a homogeneous freight corridor. This was then compared to a shared corridor setting in which 25% of the total traffic was passenger trains. In order to differentiate between delay mechanisms, this shared corridor condition was simulated twice. The first run was with low-speed, high-priority passenger trains to determine the impact of priority. The second run determined the marginal effect of a speed differential by using higher-speed, higher-priority passenger trains. Through this analysis, the capacity impact of the passenger trains on freight railway operations can be attributed to specific delay-causing mechanisms. These results can aid railway planners in determining the amount of double track needed to mitigate the effect of additional traffic on a rail corridor. The results of these experiments provide a better understanding of key fundamental relationships affecting railway performance. The results presented in this chapter are not intended to represent absolute predictive measurements for a particular set of conditions. Rather, they are meant to illustrate comparative effects under different conditions.

### **7.1.1 Background**

Most railway infrastructure in the United States is owned by private freight railway companies (Bing et al. 2010). With few exceptions, public passenger train operating agencies must negotiate access to the freight railway lines to provide passenger service. In most circumstances, the goal

is that the level of service of the freight railway should be unaffected by the additional passenger traffic added to the route. This is usually accomplished by installing more track to mitigate potential delays. In some situations, the passenger agency may pay freight railroads for the slots that the passenger trains consume.

Experience and previous research have shown that double and single-track railway lines behave differently (Abril et al. 2008). Single-track railways have considerably lower capacity than double-track lines. The primary reason for this reduction in capacity is due to trains traveling in opposite directions having to take turns using the single-track sections between passing sidings. These single-track sections are often bottlenecks that constrain overall line capacity. On double-track mainlines, theoretical capacity is primarily affected by the following distance between trains moving in the same direction. Double-track capacity is reduced if there are speed differentials, overtakes, or trains traveling against the current of traffic on the opposite direction track (Kahn 1979).

Previous studies of single track showed that adding a high-priority train to a freight network will increase average train delays more than simply adding another train that has similar characteristics to others operating on the line. Vromans et al. (2006) used simulation analyses to investigate strategies to improve passenger operations. Leilich (1998) and Dingler et al. (2009) used simulation analysis to evaluate the interactions between unit freight trains and high-priority container trains, and found a capacity loss due to the heterogeneous operations. Sogin et al. (2011) simulated single-track shared corridor conditions and determined that the maximum speed of the high-priority train had little relationship with train delay. Their analysis showed that delay distributions skewed to the right with none of the trains performing close to the minimum run time. Meets at sidings were cited as a primary delay mechanism. Double-track configurations

are more sensitive to speed differentials than on single track. A faster, high-priority train may need to use the second track in the opposing direction to overtake a slower train, thereby interfering with oncoming traffic on that track. Double-track configurations have delay distributions similar to exponential distributions with many trains operating close to the minimum run time (Sogin et al. 2012a).

There has been limited research on hybrid-track configurations. Petersen et al. (1987) used simulation analysis to locate longer sidings in order to accommodate passenger trains on a freight line. Additionally, Pawar (2011) used analytical models to determine the length of long-sidings in order to run a single-track, high speed railway without delays in meets. Lindfeld (2012b) compared partial double track to additional sidings and determined that it offers more flexibility to timetables and improves practical capacity more than additional sidings.

Regression modeling of train delays has been used to quantify the effects of various operational factors on train delay. Prokopy and Rubin (1975) used single-track simulation results to develop a multivariate regression model. Kruger (1999) developed a similar approach using an updated simulation model and summarized his data with multivariate regression. Both models were developed by varying one parameter at a time to isolate the effects of each. Mitra et al. (2010) developed an 8-variable regression model using simulation results for single-track lines, but their model did not consider interaction effects between variables. Lai and Huang (2012b) used regression and neural networks to model Rail Traffic Controller (RTC) simulation results from both a single and double-track network. For both the single and double-track models, Lai and Huang used a full-factorial experimental design that analyzed five factors at three different levels.

### 7.1.2 Delay as a Proxy for Capacity

Measuring railway capacity is non-trivial. Theoretical capacity can be measured using analytical techniques; however, when measuring practical capacity it can be difficult to incorporate all the stochastic factors affecting variation in train operations. In the United States, it is common practice to simulate current railway traffic and then re-execute the simulation with additional traffic on the existing infrastructure. The differences in delay between these two cases are analyzed and in most cases, train delays increase. These delays can be mitigated by constructing additional trackage. A series of alternative infrastructure configurations is then simulated, and the one that yields the best return on investment is generally selected for construction. This process does not usually include explicit calculation of practical capacity.

Another approach to determining railway capacity is analysis of railway delay-volume curves. Using this method, train delay can be predicted as an exponential function of traffic volume, (Equation 7-1) (Krueger 1999, Mattsson 2007, Lai & Huang 2012b):

$$D = Ae^{kV} \quad (7-1)$$

Where:

$D$  = Average train delay

$A$  = Delay mitigation constant

$k$  = Congestion factor

A railway could then define the capacity of the line as the traffic volume (number of trains per day) where the level of service (LOS) deteriorates to a minimum level of service (MLOS) that is still acceptable. The exact definition of MLOS differs depending on the infrastructure owner and railway operator. For my thesis, MLOS will be defined as the maximum



average train delay that is tolerable to the railway operator,  $D_{max}$ . Under this definition for MLOS, Equation 7-1 can then be rearranged and solved for railway capacity.

$$V = \frac{1}{k} \ln\left(\frac{D_{max}}{A}\right) \quad (7-2)$$

Consider two different single-track routes that have different amounts of double-track sections installed. The delays from these two routes can be explained by Equation 7-1. Assuming that each route is operating at traffic volume at the MLOS,  $D_{max}$ . Then using Equation 7-3a, the difference in capacities can be solved. The capacity difference of the capacities of these two lines is independent of the MLOS.

$$V_2 - V_1 = \frac{1}{k_2} \ln\left(\frac{D_{max}}{A_2}\right) - \frac{1}{k_1} \ln\left(\frac{D_{max}}{A_1}\right) \quad (7-3a)$$

$$V_2 - V_1 = \left(\frac{1}{k_1} + \Delta\right) \ln\left(\frac{D_{max}}{A_2}\right) - \frac{1}{k_1} \ln\left(\frac{D_{max}}{A_1}\right) \quad (7-3b)$$

$$V_2 - V_1 = \frac{1}{k_1} \ln\left(\frac{A_1}{A_2}\right) + \Delta V_2 \quad (7-3c)$$

If the delay growth constants between the two routes are approximately equal then  $k$  will be approximately equal to 0. In this case, Equation 7-3c can be approximated by Equation 7-4. This may be a reasonable assumption when the types of traffic interactions between two infrastructure configurations are similar. However, a homogenous freight line may have a growth constant that is different with mixed passenger and freight traffic. The change in capacity described by Equation 7-4 is independent of any delay standard that a railway might set. Equation 7-4 could be used as a base for comparing the capacity improvement by adding sections of double track if the three coefficients can be related to the amount of double track installed.

$$V_2 - V_1 = \frac{1}{k} \ln \left( \frac{A_1}{A_2} \right) \quad (7-4)$$

I presume that there is some functional relationship describing the relationship between capacity and the percentage of double track. Five hypothetical transition functions are shown in Figure 7-1. These five curves are all upward sloping assuming that capacity will increase as more track is added. If capacity were measured on the y-axis, then there would be an assumed positive relationship with the amount of double track installed. The shape of these curves may be different for different performance metrics. In this chapter, I will try and identify the functional relationship for train delay and capacity under various transition scenarios from single to double track. The shapes of these curves may differ for different performance metrics.

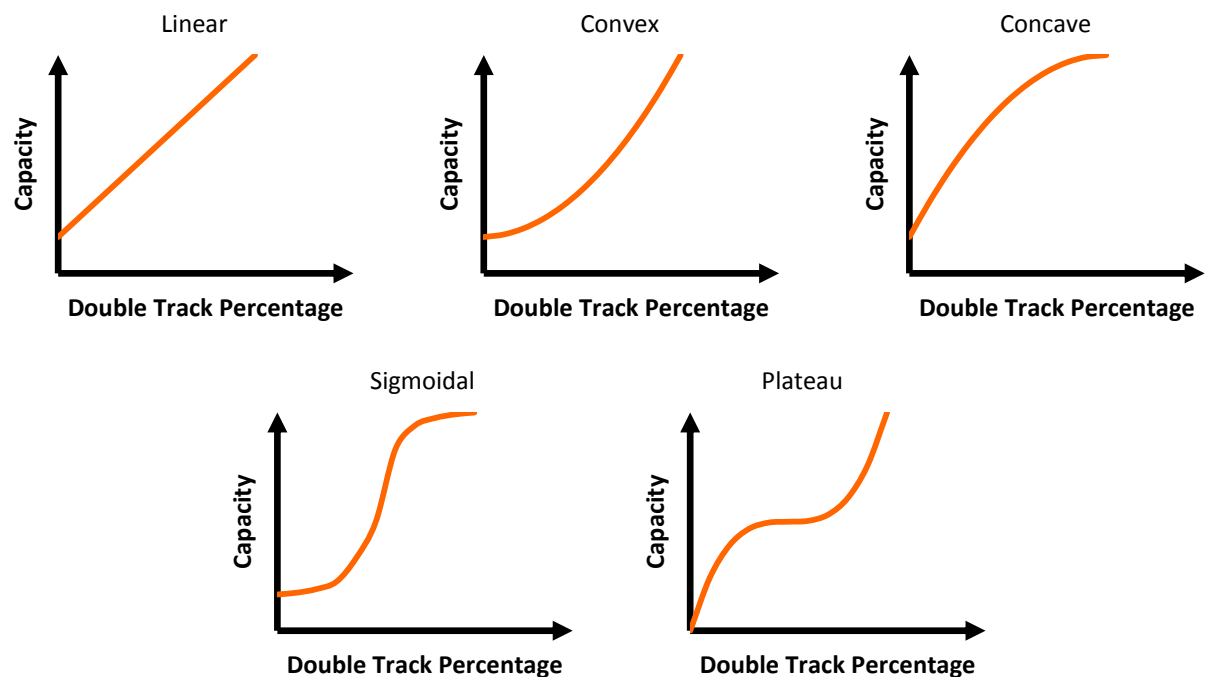


Figure 7-1: Potential shapes of transition functions from single to double track.

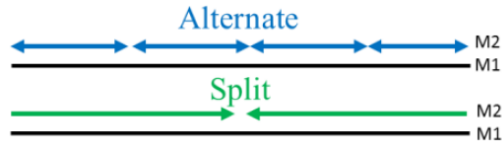
## 7.2 Simulation Methodology

The hybrid track configuration experiment examined traffic volume, traffic mixture and the amount of second-mainline track installed; and used train delay as the response variable. The original single-track line parameters are summarized in Table 7-1. This baseline is typical of a high quality, single-track mainline in North America with high turnout entry speeds, dense siding spacing, and capacity for long trains.

There are various strategies for how the second mainline track could be constructed in phases and distributed across a corridor. There will likely be sections of the route that cost more to construct than others. Based on a strategy of minimizing capital investment per mile of double track, the inexpensive sections of double track would be expected to be constructed before the expensive sections. However, if cost was not a factor in the decision making process, or if the cost of all double-track segments was relatively consistent, then there might be an optimal distribution of double track from an operating perspective. Two potential grouping strategies are illustrated in Figure 7-2: *alternate* and *split*.

**Table 7-1: Route Parameter Guidelines.**

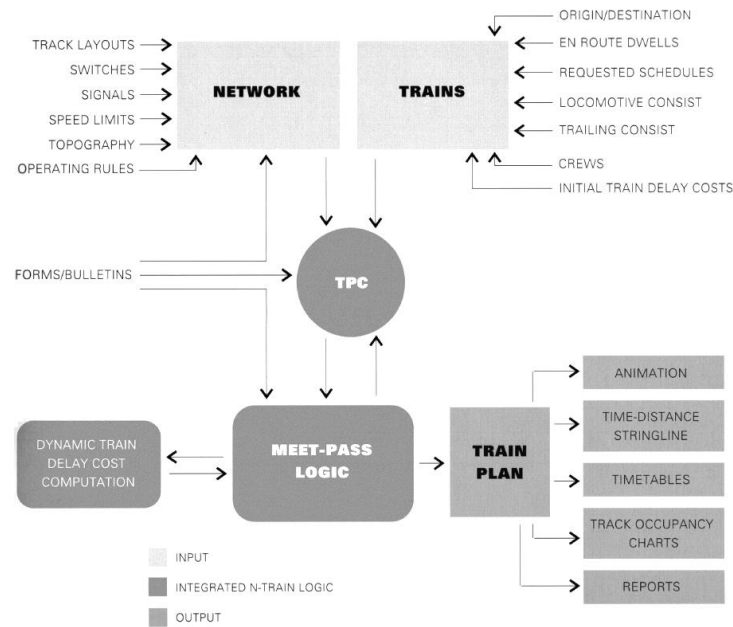
Parameter	Value
Length	240 miles
Bottleneck Length	8 miles
Siding Spacing	10 miles
Siding length	10,000 feet
Diverging turnout speed	45 mph
Traffic control system	2-Block, 3-Aspect CTC
Average signal spacing	2 miles



**Figure 7-2: Double-track allocation Strategies.**

The *alternate* strategy is to pick four to six points on the line and build-out in both directions from these midpoints. This has the benefit of creating long sections of double track where two trains can meet without either train having to stop at sidings. Additionally, these double-track sections may be long enough to achieve an overtake maneuver, where a faster train overtakes a slower train by using the opposite track. Another hybrid strategy is to *split* the double-track resource between the two terminals on the line and build towards the midpoint. The *split* strategy has the benefit of addressing potential bottleneck constraints at terminals. This provides longer double-track sections than the alternate condition with the trade-off being a longer section of single track in the middle of the route between the double-track sections at the ends.

In both allocation strategies, double track is being added to connect pairs of pre-existing sidings. When a siding becomes part of a section of mainline track, its track speed is upgraded to match the rest of the mainline. The amount of second track installed is described by the double-track percentage. For the purposes of this chapter, the double-track percentage includes both the length of sidings and second-mainline track. In this case, the baseline single-track configuration had 45.6-miles of passing of passing sidings and therefore classified as 19% double track. This accounting of double-track percentage is the same as the ratio of double-track miles per route mile.



**Figure 7-3: Architecture of Rail Traffic Controller (Wilson 2012).**

RTC, developed by Eric Wilson of Berkeley Simulation Software, is the *de facto* standard for railway simulation analyses in the United States (Wilson 2012). Users include all the U.S. Class I railways, Amtrak, the Surface Transportation Board, Bay Area Rapid Transit, major consulting firms and many others. RTC calculates train movements over a route taking into account allowable track-speeds, grade, curvature, and signaling systems. RTC will also modify train paths when trains are in conflict with each other, such as two trains simultaneously requesting use of the same section of track. Once a conflict is recognized, the logic reroutes and/or delays trains, as needed according to the priority of the trains. Trains are initially assigned user-defined priorities and departure times. As conflicts are resolved, train priorities are varied dynamically within user-defined bounds. The priority of a train varies as a function of its on-time-performance. For example, late trains are given priority over early trains. Additionally, the priorities can be adapted to reflect business objectives such as giving preference to container

trains over bulk commodity trains or passenger trains over freight trains. The architecture behind RTC is shown in Figure 7-3 (Wilson 2012).

An RTC simulation run for a particular scenario analyzes five days of traffic in the corridor. Each simulation run is then repeated six times to yield performance statistics for 30 days of railway traffic for each of the runs in the experimental matrix. If a particular randomized run was infeasible for RTC to dispatch, then it is likely that one or more of the other six replicates was feasible and the scenario can still be used in the final analysis. Additionally, replication gave the opportunity for the dispatching algorithm to make different decisions with similar inputs to the model.

Although the infrastructure differs between cases to reflect the varying amount of double track, the boundary conditions are kept constant. The route features only one origin-destination pair and traffic is directionally balanced. Each end of the route features terminals designed to minimize terminal-mainline interference by having long leads, and excess receiving and departure tracks. The double track is installed to minimize reconfiguring turnouts and control points of the signaling system. When a section of double-track connects a siding as mainline track, the turnout for that siding is reused as part of a future universal crossover leading into a future section of double track. The result is crossovers that are spaced at approximately 10-mile intervals on the double-track segments.

In addition to the physical and operating characteristics identified in Table 7-1, the schedule of traffic affects corridor performance. In many railway operations, conflicts between trains are carefully planned in a timetable. However, railway operations in North America are highly variable due to fluctuations in freight traffic demands, weather, and other sources of variation and delay. Dispatchers are resolving conflicts between trains in real time instead of

following a strict timetable. Different schedules of trains for the same infrastructure can show different average train delays. Consequently, assuming only one schedule for any given infrastructure may result in large experimental error, which will lead to bias in the results. In order to counteract error due to schedule bias, the departure time of each train is determined using a random uniform distribution over a 24-hour period. The randomized scheduling process is expected to create a range of schedules such that “good”, low-delay schedules are averaged against “bad”, high-delay schedules over the set of simulation replicates for each scenario. Stable averages can be achieved by averaging train performance over multiple days.

Each train in the simulation is based on the characteristics specified in Table 7-2. The freight train characteristics are based on the Cambridge Systematics National Rail Freight Infrastructure Capacity (2007) and Investment Study conducted for the Association of American Railroads (AAR). Freight car tonnages and lengths are based on averages for each car type. The power-to-ton ratios are based on experience and information from the Transportation Research Board Workshop on Railroad Capacity and Corridor Planning (Galloway 2002). The passenger train is based on the 110-mph operation between Chicago and Detroit. The acceleration curves from the simulation model were matched to GPS coordinate data from the Amtrak geometry car on this corridor. The passenger train stops were spaced at 32.4 mile intervals based on the current Amtrak station spacing on routes in California, Illinois, Washington, and Wisconsin. A 50-mph maximum freight train speed is typical for a well maintained, high-density mainline. Without incab signaling systems, the maximum speed of passenger trains is limited by regulation to 79-mph in the U.S. (Federal Railroad Administration 2013c). Potential maximum speed upgrades on developing shared corridors across North America are 90-mph and 110-mph.

**Table 7-2: Train Parameters for Simulation Model.**

	Freight Train	Passenger Train
Locomotives	x3 SD70	x2 P42
No. of Cars	115 hopper cars	8 Single level cars
Length (ft)	6,325	740
Weight(tons)	16,445	800
HP/TT	0.78	15.4
		32.4 mi between stops

Train delay is used as a proxy for capacity in this chapter. While train delay is not a flow measurement, large train delays are indicative of a congested or saturated network. Train delays can be predicted by the traffic volume that can then give insight into the capacity of the railway line. Passenger traffic has a lower tolerance for delay compared to most freight traffic. These analyses will focus on freight train delays because these delays are most responsive to the identified factors. Passenger trains are shielded from experiencing very high delays due to their higher priority. Regression models of passenger train delays will be developed in the future.

Previous work on shared corridors has identified two major characteristics of passenger trains that may cause delays to freight trains: higher priority and speed differentials (Sogin et al. 2011, 2012b). The effect of priority was analyzed by having a high priority passenger train travel at the same speed as the freight trains, 50-mph. The effect of speed differential and priority acting together was represented by 110-mph high priority passenger trains and 50-mph low priority freight trains. In the context of the simulation software, priority is a measure of preference. There may be situations where delaying a passenger train can result in better network fluidity. For example, an eastbound passenger train may stop on a siding in a meet with two oncoming westbound freight trains. This can result in lower network delays than by splitting this conflict into two separate conflicts at two sidings. In single-track networks, passenger trains are



often delayed by meets with other high priority passenger trains traveling in opposite directions because the siding length is only two miles.

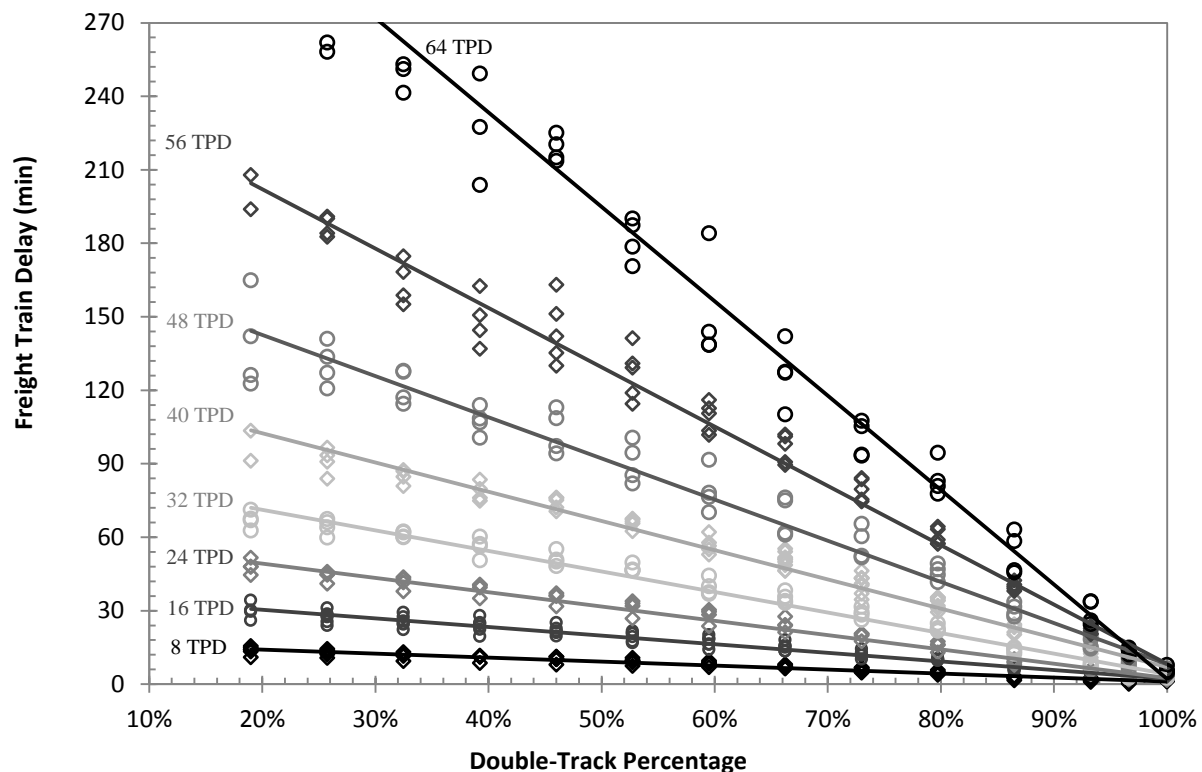
### 7.3 Developing the Response Surface Model

The following analysis will focus on developing a response surface model based on the simulation data. The goal is to be able to predict the capacity of a line as a function of the amount of double track installed and the MLOS. The analysis in this section will show the development of a response model for a freight-only corridor where the double track is allocated in an *alternate* strategy.

The evidence from this study suggests that train delay will decrease linearly for each marginal section of double track added to the single-track baseline condition using an *alternate* strategy. This linear decrease in train delay occurs in each of the eight different traffic levels studied from 8 trains per day (TPD) up to 64 TPD. Figure 7-4 shows the freight train delays at the eight different traffic levels over 14 different track configurations progressing from pure single track (19-percent) to complete-two-mainline track (100-percent). The linear reduction in train delay is greater with higher traffic levels than with lower traffic levels. Additionally, these trend lines are pivoting around approximately 100% double track.

Each of the trend lines in Figure 7-4 can be described by slope and intercept parameters. If there are clear relationships between these parameters and the traffic volume on the line, then there can be a master equation that predicts delay for a given double track percentage and traffic volume. This equation would be in the form of Equation 7-5. The intercept  $\gamma_0(V)$  and slope  $\gamma_1(V)$  are both functions of volume. In an alternate formulation, the slope term could represent the delay reduction per mile of double-track installed under constant volume. In this analysis, the

double-track parameter is normalized by route length so the slope parameter is the reduction in train delay per double-track-percentage point. An important property of Equation 7-5 is that it is centered on 19-percent, the single-track configuration. This point-slope format results in the intercept term relating to the amount of train delay in a single-track configuration. Otherwise, in slope-intercept format, the y-axis intercept would indicate a theoretical amount of delay on single track with zero sidings. Values in this range were not simulated and violate the route parameter guidelines in Table 7-1. The x-axis intercept is an indicator of the level of double tracking where the line experiences no train delays. In the cases simulated, this value was greater than 100% and indicates a small amount of triple-track.



**Figure 7-4: Train delays as a function of percentage double track at various traffic volumes.**

$$D(V, x) = \gamma_0(V) - \gamma_1(V) \cdot (x - 19\%). \quad (7-5)$$

Where:

$\gamma_0(V)$  = Single-track train delay as a function of traffic volume (intercept)

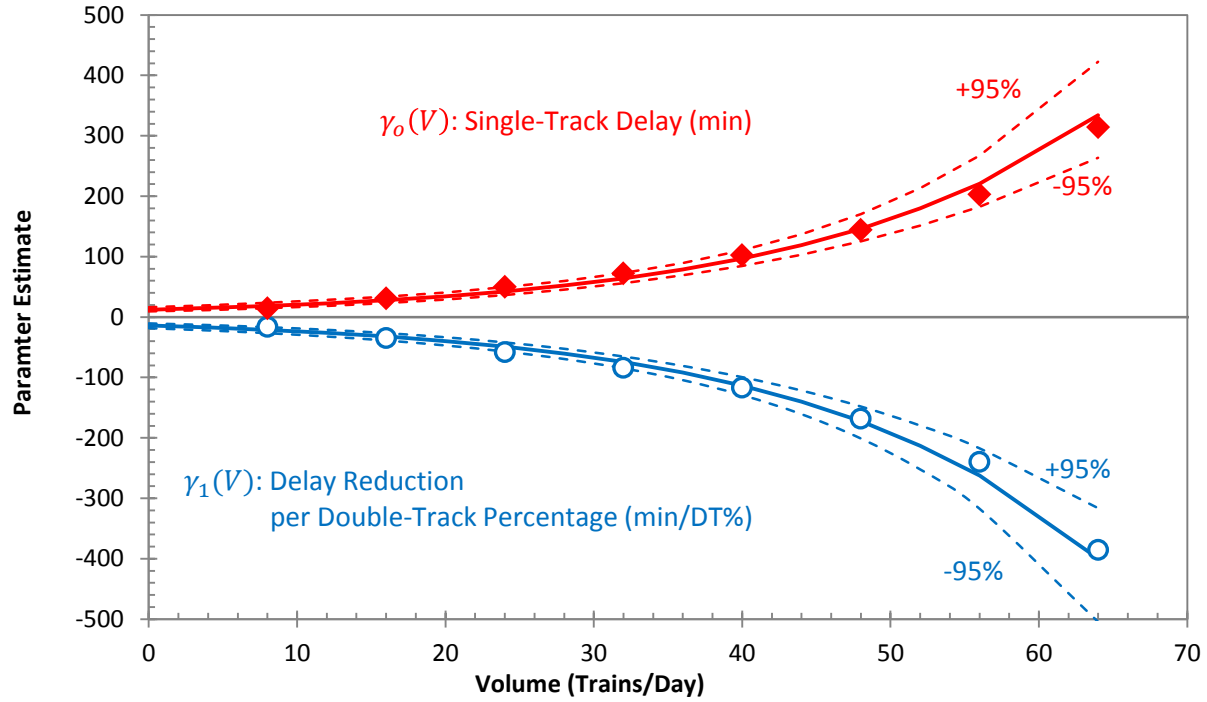
$\gamma_1(V)$  = Reduction in train delay per double track-percentage point as a function of traffic volume (slope)

$x$  = Double-track percentage

The parameters of the linear trend lines for each traffic volume are shown in Table 7-3. The single-track intercepts are always positive and increase with higher traffic volumes. The negative slope terms increase in magnitude with higher traffic volumes (Figure 7-4). These trends in opposite directions can describe the pivoting of the trend lines around approximately full double track in Figure 7-4. Both the slope and single-track-delay parameter are plotted as points against volume in Figure 7-5. The relationship between these trend-line parameters and traffic volume can be explained by several different functional relationships including exponential or polynomial. An exponential relationship only requires two parameters to describe  $\gamma(V)$  whereas a polynomial would require at least three, assuming at least a quadratic fit. In higher order polynomials, there is greater difficulty in deriving physical meaning from parameter estimates. Lastly, using an exponential relationship is more likely to simplify to an equation similar to Equation 7-1. If exponential relationships are assumed then Equation 7-5 becomes Equation 7-6.

**Table 7-3: Linear Parameter Estimates of Train Delay under Constant Volume.**

Volume (TPD)	Slope, $\gamma_1(V)$ (min/Double-Track %)	Single-Track Delay, $\gamma_0(V)$ (min)
8	-16.3	14.2
16	-35.0	30.6
24	-58.4	49.8
32	-83.9	72.0
40	-117.1	102.3
48	-168.1	144.2
56	-239.9	203.1
64	-385.5	314.4



**Figure 7-5: Linear parameter estimates of train delay reductions by adding sections of two- mainline-track. These parameter estimates are then predicted by exponential relationships.**

$$D(V, x) = A_o e^{k_o V} - A_1 e^{k_1 V} \cdot (x - 19\%) \quad (7-6)$$

Where:

$A_o$  = Single track train delay constant

$k_o$  = Single track congestion factor

$A_1$  = Slope constant

$k_2$  = Slope congestion factor

The solid lines in Figure 7-5 predict the linear parameters using exponential relationships. The single-track-intercept parameter can be predicted using Equation 7-7a and the slope by Equation 7-7b. These relationships were determined using a log-transformation and simple linear regression procedures in JMP (JMP 2013). The dashed lines represent 95% confidence bands around the mean response of the linear parameter estimate. Equations 7-7a and 7-7b can accurately predict the linear parameter estimates and be substituted into Equation 7-6 and yield Equation 7-8.

$$\gamma_0(V) = 12.27e^{0.05162V} \quad (7-7a)$$

$$\gamma_1(V) = 13.89e^{0.05249V} \quad (7-7b)$$

$$D(V, x) = 12.269e^{0.05162V} - (13.889e^{0.05249V})(x - 19\%) \quad (7-8)$$

A disadvantage in the method of producing Equation 7-8 is that each time the data are passed through a simple linear regression, degrees of freedom are lost. In this case, eight linear trend lines were determined, each featuring two parameter estimates. Equations 7-7a and 7-7b also require two parameter estimates. In total, 20 different parameter estimates were determined to derive Equation 7-8. As an alternative to this hierarchical regression approach, non-linear regression can be used to arrive at the final four parameter estimates of Equation 7-6 without the loss in degrees of freedom. Not surprisingly, the non-linear regression platform yields more precise parameter estimates and results in a lower root mean square error (RMSE) of the original data.

**Table 7-4: Comparison of the Hierarchical & Non-Linear Regression.**

			Statistic	Hierarchical	Non-Linear
			RMSE	31.04	8.29
Single-Track Delay Intercept	$A_0$	Estimate		12.269	13.201
		Lower Limit		9.237	12.554
		Upper Limit		16.298	13.881
	$k_0$	Estimate		0.05162	0.0495
		Lower Limit		0.04459	0.0486
		Upper Limit		0.05864	0.0504
Slope (delay reduction)	$A_1$	Estimate		13.889	14.345
		Lower Limit		10.478	13.197
		Upper Limit		18.410	15.594
	$k_1$	Estimate		0.05249	0.0513
		Lower Limit		0.04551	0.0498
		Upper Limit		0.05947	0.0527

An important aspect of the parameter estimates for Equation 7-6 in Table 7-4 is the similarity between  $A_0$  and  $A_1$  as well as the estimates for  $k_0$  and  $k_1$ . With 95% confidence intervals, there is clear overlap between these pairs of parameter estimates. While the model of Equation 7-8 is significant and built on sound theory, a simpler model may be sufficient. In

particular, if  $k_0$  and  $k_1$  are equal, then the freight train delay data can be described by Equation 7-9. This equation is no longer centered on 19% double track for the purposes of simplicity. Equation 7-9 is in the form of Equation 7-1 where the A term is now described by a linear function of the double-track percentage. An interesting property of Equation 7-9 is that it has a closed-form solution when solved for traffic volume instead of delay as shown in Equation 7-10. In this form, the capacity of the line is then a function of the amount of double track installed and a delay standard,  $D_{max}$ .

$$D = (S_1 - S_2x)e^{kV} \quad (7-9)$$

where

$S_1$  = Single-track delay constant

$S_2$  = Delay mitigation constant

$k$  = Congestion factor

$$V = \frac{1}{k} \cdot \ln \left( \frac{D_{max}}{S_1 - S_2x} \right) \quad (7-10)$$

Equation 7-10 is plotted in Figure 7-6 for different delay standards. The capacity improvement of the double-tracking is close to linear when the line is closer to a single track-configuration. As more double track is added and the line approaches full two main track, the additional segments of second track yield increasingly greater capacity benefits. These capacity curves can help justify the last mile investments to complete the double-tracking of a line. These may be expensive tunnels, bridges, mountain passes or improvements in urban areas. The delay standard has more of an effect of determining the capacity of the line when the standard is low. At higher delay standards, the capacity contours are grouped much more closely (Figure 7-6).

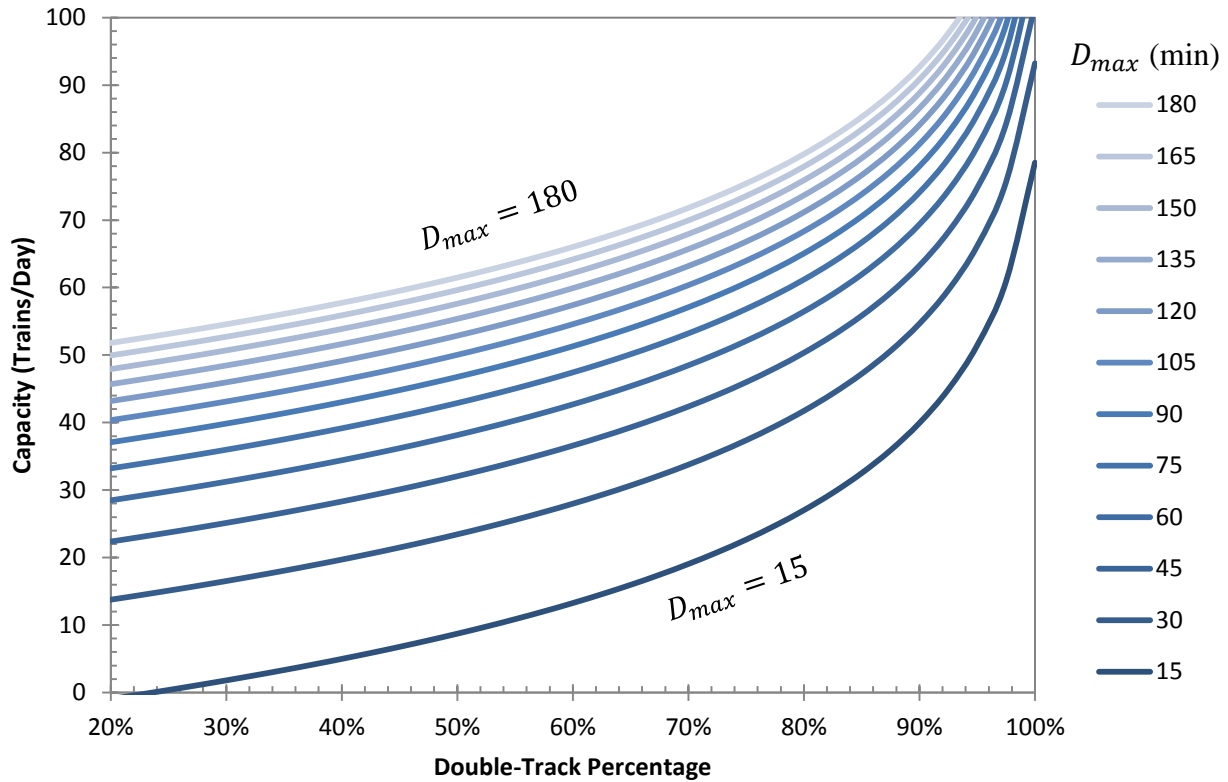


Figure 7-6: Capacity as a function of the amount of double track installed under different delay standards,  $D_{max}$ .

The instantaneous slopes of the curves plotted in Figure 7-6 are parallel at all double-track percentages. This property is verified by taking the partial derivative of Equation 7-10 with respect to the double-track percentage, which yields Equation 7-11. The implication of this is that the change in capacity from installing sections of double track is independent of the delay standard,  $D_{max}$ . For example, capacity increases by 5.8 TPD at each delay standard when upgrading from 60% double track to 70% double track. Because Equation 7-11 does not include the delay standard,  $D_{max}$ , changing the delay standard will result in the in a constant change in capacity that is independent of the double-track delay percentage,  $x$ . For example, changing the

delay standard from 30-minutes to 60-minutes will increase capacity by 14.2 TPD regardless of double-track percentage.

$$\frac{\partial V}{\partial x} = \frac{S_2}{k(S_1 - S_2 x)} \quad (7-11)$$

#### 7.4 Comparing Different Operating Conditions

The previous analyses were completed using only homogenous freight trains and only one method of allocating sections of mainline track to a railway corridor. In the following analyses, the effect of additional parameters was considered using Equations 7-9 and 7-10. First, the double-track allocation strategy will be changed from *alternate* to *split* (Figure 7-2). Instead of simulating eight different traffic levels for all infrastructure configurations, only traffic levels of 16, 40, and 56 trains per day will be considered in order to develop parameter estimates for Equation 7-9 for other conditions. The capacity of two or more configurations is compared by using Equation 7-10 at a MLOS set to 60-minute average delay. In the previous section, the change in capacity was independent of the MLOS. In order to use Equation 7-10 to compare different operating conditions, six different parameters must be estimated and the MLOS does not drop out of the model. Fortunately, the partial derivative of volume with respect to double-track percentage is much greater than the partial derivative of volume with respect to  $D_{max}$ , (Equation 7-12). In the case of the freight-only corridor at 60% double track and a 60-minute MLOS, using an alternate strategy, a unit change in double track is greater than a unit change in the delay standard by a factor of 143. As long as the parameter estimates of Equation 7-9 are of the same magnitude as those estimated in the previous section, then the MLOS will have a small effect on the change in capacity between different operating conditions.

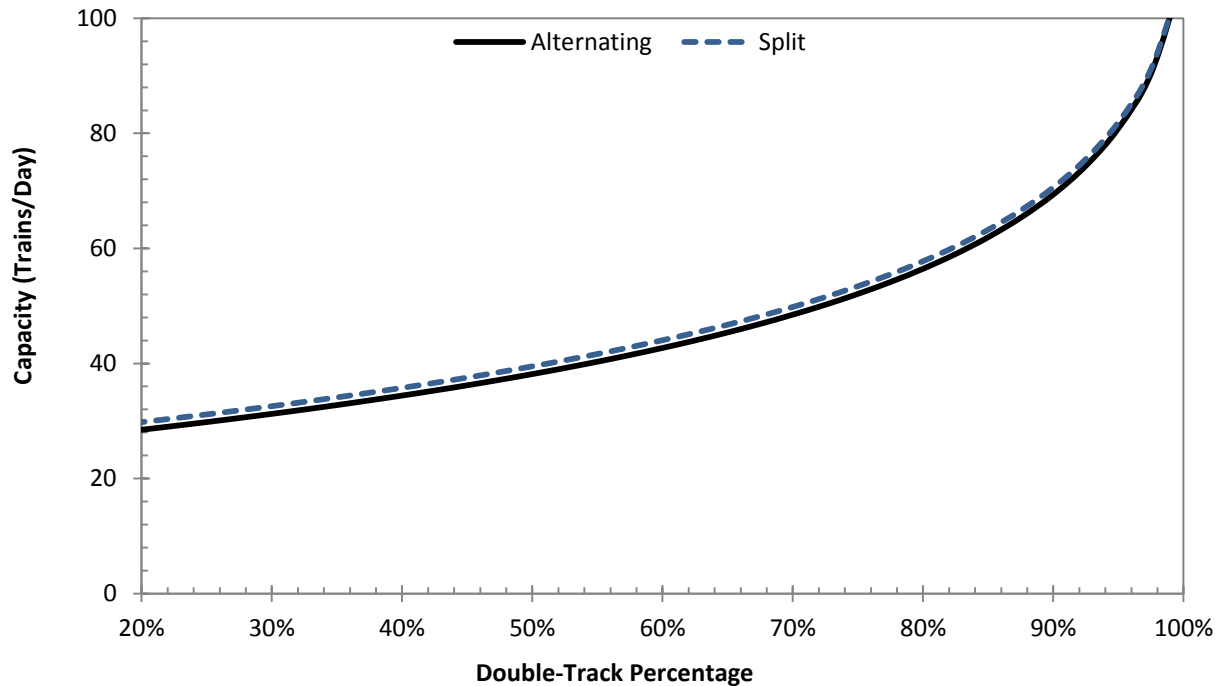


$$\frac{\partial V}{\partial D_{max}} = k(D_{max})^{-1} \quad (7-12)$$

The difference between these two allocation strategies for homogenous freight trains does not lead to significantly different changes in line capacity. Allocating track in a *split* configuration instead of an *alternate* configuration shows an improvement of about one-half train per day (Figure 7-7). The parameters for these two scenarios are summarized in Table 7-5. While there may not be much change between the two infrastructure configurations under the operating conditions of this analysis, there may be greater changes in the results under different scenarios. For example, if more sophisticated terminal effects were included in the model then the split configurations may show more benefit. If the model followed a strict timetable, then meets between trains can be planned to occur at sections of double track and take better advantage of the *alternate* configuration.

**Table 7-5: Parameter Estimates for Different Allocation Strategies.**

	Alternate	Split	% Change
$S_1$	19.5206	18.4404	-5.53%
$S_2$	19.1490	18.0585	-5.69%
$k$	0.0471	0.0469	-0.42%



**Figure 7-7: Capacity improvement under a 60 minute delay standard when sections of double track are allocated to the terminals (split) or at a collection of midpoints along the line (alternate).**

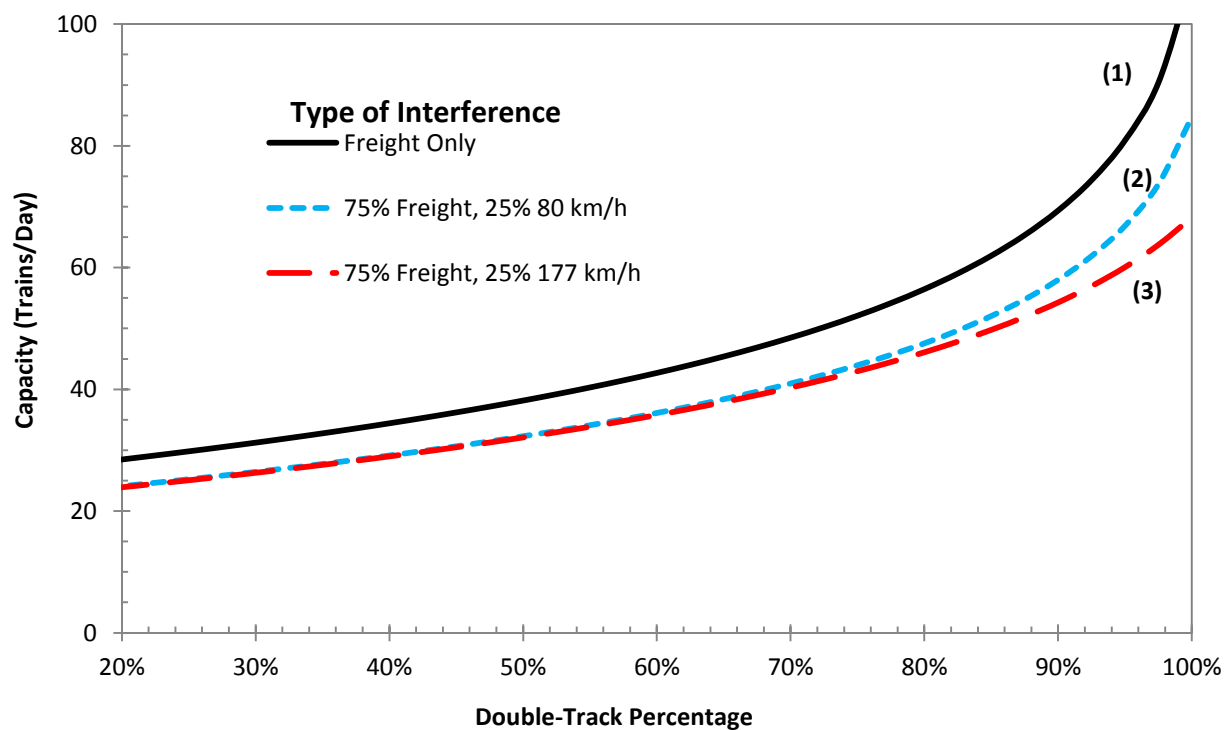
I considered two potential mechanisms that might cause additional delay to freight trains from passenger trains on lines where they share trackage. The first is a priority differential between train types in which passenger trains are given preference in meet or pass conflicts. The second delay mechanism studied was speed differentials between train types. The effect of priority will be illustrated in a mixed traffic line where there are three freight trains per passenger train. Both trains will be limited to a maximum speed of 50-mph. The effect of speed and priority was evaluated by having 25% of the total traffic comprised of 110-mph passenger trains. The double track was allocated using the *alternate* strategy.

A heterogeneous mixture of three freight trains per passenger train will by itself result in a capacity loss for any delay standard. Additionally, it will take more double-track to mitigate traffic increases. Having priority trains only manifests a change relative to the base case by

having a higher  $k$  coefficient, indicating higher sensitivity to traffic increases. The higher speed passenger trains have a  $k$  coefficient on par with the freight-only case but also have higher  $S_1$  and  $S_2$  coefficients (Table 7-9). The capacity curves for 60-minute MLOS are plotted in Figure 7-8. The difference between 50-mph and 110-mph is very small between 20% and 80% double track. In the 80 to 100% double-track range, there is a divergence between the two passenger train interference curves, where speed differentials start to reduce capacity.

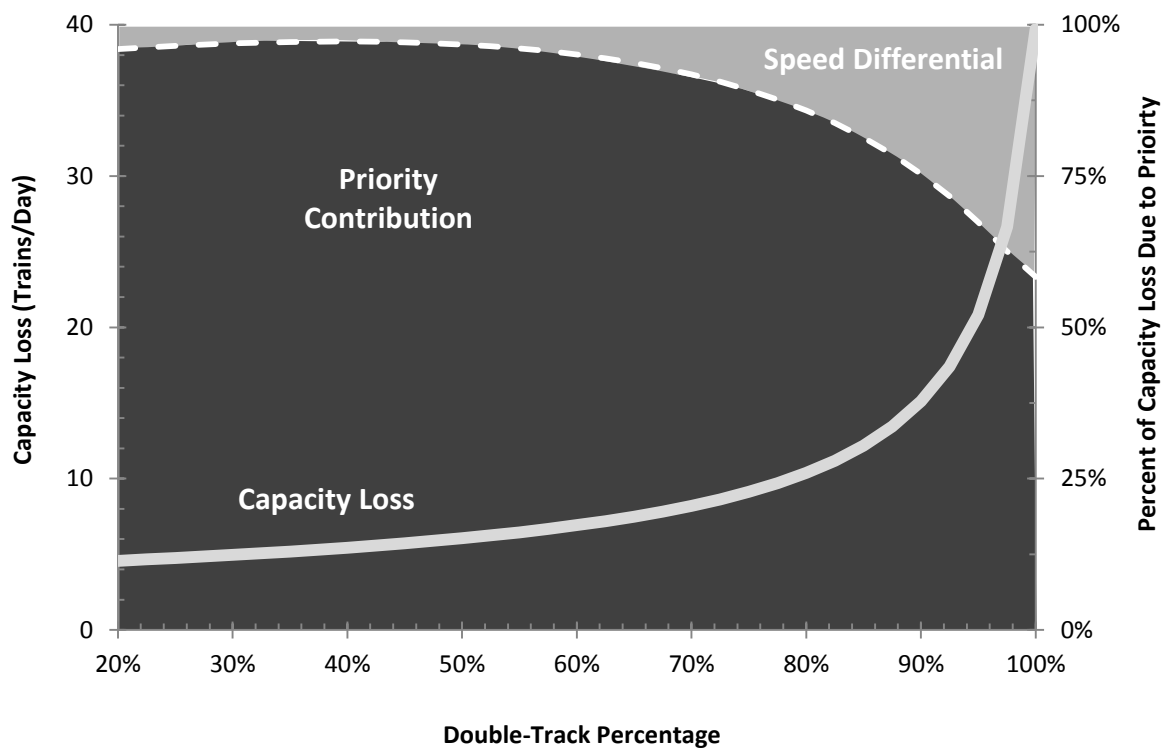
**Table 7-6: Parameter Estimates of Equation 7-9 for Different Traffic Types.**

	Freight Only	50 mph Passenger Trains	110 mph Passenger Trains
$S_1$	19.5206	19.9317	22.4534
$S_2$	19.149	19.3509	20.4052
$k$	0.0471	0.0547	0.0495



**Figure 7-8: Change in capacity by installing sections of double track with a 60-minute delay standard. Where (1) the traffic is 100% freight trains, (2) the traffic is 75% 50-mph freight trains and 25% 50-mph passenger trains, and (3) the traffic 75% 50-mph freight trains and 25% 110-mph passenger trains.**

The loss in capacity due to 110-mph passenger trains in a freight corridor under an average 60-minute MLOS is illustrated in Figure 7-9. The capacity loss due the higher speed passenger train is greatest on double track and lowest on single track. Additionally, this loss curve has a convex transition where the change in capacity is greater at higher double-track percentages than at lower percentages. The dashed curve of Figure 7-9 estimates the delay mechanism by comparing the capacity loss due to priority with 50-mph passenger trains, to the total loss from 110-mph passenger train interference. The effect of priority differential accounts for 96% of the capacity loss between 19-and-55% double track. At full 100% double track, the speed differential mechanism accounts for 41.7% of the loss in capacity, and priority accounts for 58.3-percent.

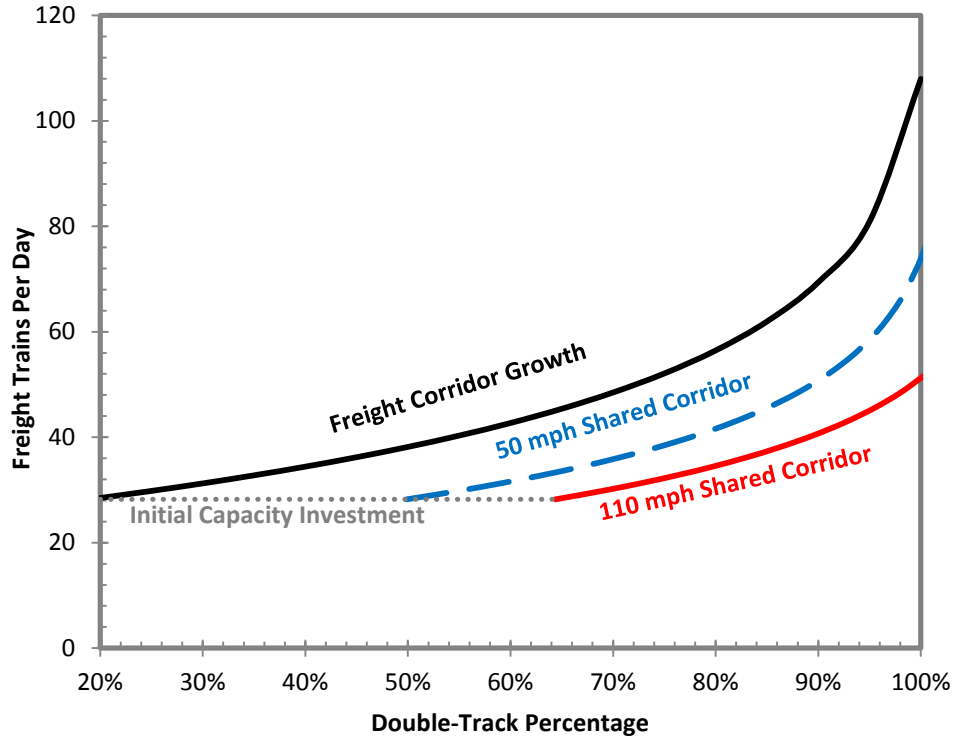


**Figure 7-9: Capacity loss of operating higher-speed passenger trains and the estimated delay-mechanism contribution.**

The freight train capacity loss on a higher-speed passenger, mixed use corridor is even greater than that shown in Figure 7-9 because it assumes 25% of the available capacity is being used to accommodate passenger trains instead of freight trains. Consider a case where a railway line is originally a single-track, freight-only corridor at full capacity. The long term plan is to change this line into a mixed-use corridor where future traffic is comprised of 75% freight and 25% 110-mph passenger trains. The initial capital investment mitigates the additional delays to the original freight trains and improves capacity by 33 percent to accommodate the additional passenger trains. If the freight line was originally single track, then this initial investment will be to upgrade the line from 19% double track to 65% double track under a 60-minute MLOS. The amount of double track needed to accommodate the passenger trains was calculated using Equation 7-13. This substantial investment in capacity does not benefit the freight railway; it simply allows them to maintain their current level of service. In this 75-percent-freight, mixed-use corridor, any future growth in both freight and passenger volume will require additional capacity investment with 65% double track as the new baseline (Figure 7-10). If there were significant engineering cost constraints to expanding this corridor beyond two-main tracks (i.e. triple track), then the freight railway has lost its ability to accommodate new freight business in the future. If the speed differential between train types were eliminated then the initial capacity investment would be to upgrade the corridor to 50% double track. With 50-mph passenger trains, there is also more freight capacity available in a full two-track build out than with 110-mph passenger trains.

Fortunately for the freight railway, there are some short-term benefits. By having a passenger agency make the initial investments to accommodate the passenger trains, then the next time freight demand increases, the freight railway will receive a higher return on capacity

per track-mile installed. This benefit occurs because the freight railway is now on the 110-mph shared corridor curve in Figure 7-10, which is steeper at 65% double track than a freight only corridor at 19% double track.



**Figure 7-10: Freight train capacity grown under a 60 minute MLOS for a freight corridor and a shared corridor comprised of 75% freight and 25% passenger trains.**

$$x_m = S_{2p}^{-1} \left( S_{1p} - (D_{max}) e^{-V_o k_p \varphi^{-1}} \right) \quad (13)$$

where

$x_m$  = Level of double tracking to mitigate addition of passenger service

$S_{1p}$  = Shared corridor single-track delay constant

$S_{2p}$  = Shared corridor delay mitigation constant

$k_p$  = Shared corridor congestion factor

$V_o$  = Initial freight corridor volume (capacity at  $D_{max}$ )

$\varphi$  = Freight train percentage of the total traffic of the planned shared corridor

## **7.5 Future Work**

Equation 7-9 is a powerful model for relating train delays to traffic volume and the percentages of double-track. Further manipulation of various operating conditions can provide further insight on how the parameter estimates of Equation 7-9 change. For example, the three parameter estimates could be related to various levels of passenger train speeds. The derivation of Equation 7-9 depends heavily on the linear relationship between double-track percentage and train delay under constant volume. In this analysis, the sidings were all evenly spaced at 10 miles. If there was a random distribution of distances between sidings, then these trend lines may not be linear. Further investigation into this assumption is needed, as the capacity benefits of connecting longer bottleneck sections may be disproportionally greater than connecting shorter sections. Additionally, there are other strategies on how to allocate sections of double track across a corridor beyond the ones identified in Figure 7-2. RTC has limited ability to model terminal-mainline interactions. In this analysis, the terminal was designed large enough to minimize these effects; however if larger terminal-mainline interactions were designed into the model then these effects may dictate a greater effect of the allocation strategy.

## **7.6 Conclusion**

Regression analysis is a powerful technique for comparing simulation results to determine changes in capacities of different infrastructure configurations that can yield results independent of the MLOS. Train delays decrease linearly with additional sections of double track when the volume is constant. These trend-lines can be predicted and used to develop response surface models in the typical form of exponential delay-volume relationships. As a hybrid track configuration transitions from single to double track, under a constant MLOS, the incremental

capacity gained from each section of double track added increases as more double track is added to the corridor. The simultaneous operation of freight and passenger trains on a heterogeneous line can reduce capacity and the incremental capacity gained from each section of double track. The marginal loss in capacity from heterogeneous operation is greater on lines close to full double track than hybrid track configurations that are closer to single-track lines. When large speed differentials are present between train types, the speed differential may not be a significant delay causing mechanism until most of the line is double tracked.



## CHAPTER 8

### RELIABILITY APPROACH TO VARIATION IN TRAIN DELAYS

Reliability is critical to the success of both freight and passenger operations but the degree of reliability required varies widely depending on traffic type and a variety of economic factors. In all transportation modes, there is a minimum-run-time (MRT) that will be affected by various speed restrictions and vehicle performance characteristics. In many cases, scheduling transportation services to the MRT is impractical because the MRT may not be achievable on a regular basis, or there is a high probability of long delays. In order to account for this variability, extra time, often referred to as “buffer” or “slack”, is added to the schedule. These buffer times absorb delays such that actual performance of the service meets or exceeds the expectations of the users of the transportation system. This type of accounting does not change the underlying time distribution of the transportation process. For example, a commuter can drive to work in 20 minutes when all the traffic lights are green. However, the commuter leaves 10 minutes earlier in order to account for the possibility that all the traffic lights are red. The commuter has added 10-minutes of buffer time in order to achieve a desired on-time performance (OTP). This buffer did not change the distribution of possible commute times, but it did alter the economic utility. To the commuter, the cost of arriving early is less than that of arriving late. In a railroad context, this buffer affects cycle times, equipment utilization and fleet size.

In Chapter 5, I first proposed qualitatively that single and double track configurations have different delay distributions when both configurations were operating close to practical capacity. However, the Weibull distribution is capable of mathematically describing both single and double track delay distributions. In this chapter I will introduce unique characteristics and modeling capabilities of the Weibull distributions. Example railroad analogs to definitions to the

field of failure analysis will be proposed. The last section of this Chapter complements the analyses of Chapter 6 and Chapter 7 by using logistical regression techniques to quantify changes in distributions of train delays.

## 8.1 Analogies from Failure Analysis

Two types of probability density functions (PDF) of typical delay distributions will be considered in this chapter (Figure 8-1). Both PDFs are based on the Weibull distribution as this is a robust distribution that is capable of fitting many different forms of data. “Type A” is a Weibull distribution calculated from a shape and scale parameter (Equation 8-1). The “Type B” distribution is for the special case where the shape parameter of the Weibull distribution is equal to 1.0 and the resulting distribution is known as the exponential distribution, which can be calculated from a single scale parameter (Equation 8-2a & 8-2b) (Evans et al. 2000). The classification scheme of “Type A” and “Type B” is employed in order to distinguish between the unique exponential distribution with many trains performing close to the MRT, and the more typical Weibull, which features more of a bell shape and many trains experiencing large delays. While delay is analyzed in this chapter, both of these distributions could be used to analyze runtime through a data transformation in which a threshold parameter,  $\delta$ , is added to the run time of each train.

$$F(D \leq d) = 1 - e^{-\left(\frac{d}{\eta}\right)^\beta} \quad (8-1)$$

Where:

$F(D \leq d)$  = Percent of trains that have arrived within  $d$  minutes (cumulative failure)

$d$  = Delay (time)

$\eta$  = Scale parameter of the Weibull distribution (Type A)

$\beta$  = Shape parameter of the Weibull distribution (Type A)

$$F(D \leq d) = 1 - e^{-\lambda_e d} \quad (8-2a)$$

$$\lambda_e = \frac{1}{\bar{D}} \quad (8-2b)$$

Where:

$\lambda_e$  = Rate parameter of the exponential distribution (Type B)

$\bar{D}$  = Average train delay

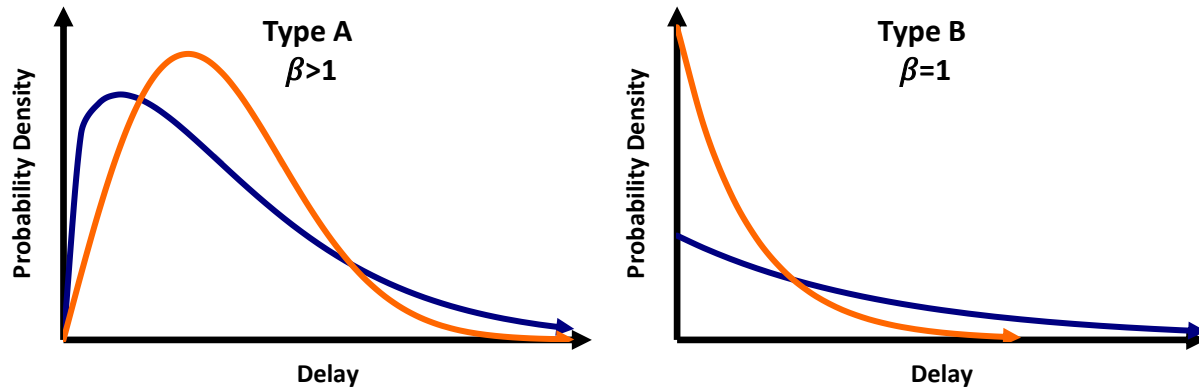


Figure 8-1: Examples of (left) Weibull and (right) exponential distribution.

The Weibull (“Type A”) distribution has been used extensively in failure and survival analyses (Washington et al. 2003). For these fields, the parameters of the Weibull distribution have physical meaning. The scale parameter,  $\eta$ , is correlated with the average useful life in the context of product failure. When the shape parameter,  $\beta$ , is greater than 1 then the product is “wearing-out” and the occurrence of a failure event increases with the age (O’Connor & Kleyner 2011). The hazard-rate represents the probability of product failure given its age. If the hazard-rate is not constant, then there can be changing expectations about the quality of the product over time. However, when the Weibull distribution is applied to train delays, these definitions from failure analysis need to be modified; trains are not “failing” by arriving late. Although typically associated with mean-time-to-failure analysis, Weibull distributions can be used in a broader context to describe probability of a certain amount of time elapsing before a particular event

occurs. Examples of non-failure related uses of the Weibull distributions are determining periods of unemployment (Kiefer 1988), time between shopping activities (Bhat 1996), time to clear highway accidents (Jones et al. 1991), and commute times (Stathopoulos & Karlaftis 2001, Al-Deek & Emam 2006). In the context of train delays, the event with the best analog to the mean-time-to-occurrence and “failure” analysis is a train arriving at its final terminal with a certain amount of accumulated delay. This accumulated delay when the train arrival event occurs is analogous to the useful life or time to failure of a particular component. Although the desired outcome is the opposite in that arrivals after a short amount of delay has accumulated are beneficial, while failures after a short amount of time has elapsed are undesirable.

Equation 8-3 is the duration-hazard rate function from failure analysis. In product failure, hazard rate measures the conditional probability at time that a failure occurs in the next instant given that the product has survived to time  $t$ . For example, the hazard rate of an automobile with 100,000 miles is greater than an automobile with 20,000 miles. The probability of the 100,000-mile car breaking down within the next mile is greater than the newer car. In the context of train delays, hazard rate measures the conditional probability density that an en route train arrives within the next minute. Similar to a PDF, the area under the curve is equal to one. To illustrate this concept, consider an intercity train that is scheduled to arrive at 12:00 and the current time is 12:15. This train is already late, and within the next minute there is a chance that the train will arrive at the terminal and a chance the train will still be en route. The hazard rate measures the conditional probability of a train arriving in the next minute given that train is already  $d$  minutes late. This duration-hazard rate is constant when the shape parameter,  $\beta$ , is 1 and increasing when  $\beta$  is greater than 1

$$S(D \leq d) = 1 - A(d) = e^{-\left(\frac{d}{\eta}\right)^\beta} \quad (8-3)$$

$$h(d) = \frac{\Pr(d < D < d + \Delta | D > d)}{\Delta} = \frac{\frac{d}{dt}[F(d)]}{S(d)} = \left(\frac{\beta}{\eta}\right) \left(\frac{d}{\eta}\right)^{-\beta-1} \quad (8-4)$$

Where:

$S(D \leq d)$  = Percentage of en route trains (survivor function)

$h(d)$  = Duration-hazard rate which equates to the conditional probability that an en route train will arrive within the next minute

A typical Weibull (“Type A”) distribution fit is shown in Figure 8-2 over a delay population of 1,280 simulated trains from Chapter 7 featuring 64 freight trains per day (TPD) and 80% double track. The bars represent the frequency distribution of train delay times dataset and the line is the Weibull function fit to the data from JMP (2013). Few trains arrive within 20 minutes of the MRT but a large cluster of trains arrive between 30 and 100 minutes late. After this interval, longer delays become less frequent. In this case, the Weibull shape parameter,  $\beta$ , was 1.62 and therefore indicates an increasing duration-hazard rate. This rate measures the instantaneous conditional probability that a train will arrive within the next instant given that it has accumulated a certain amount of delay and has not arrived yet. For example, 100 trains arrived less than 20 minutes late. Of the remaining 1,180 trains that are more than 20 minutes late, 192 are expected to terminate within next 20-minute interval. When an en route train was 20 minutes late, the conditional probability that the train will arrive with 20 to 40 minutes of delay was then 0.16 ( $192/1,180 = 0.16$ ). The duration-hazard rate is then calculated as 0.008 ( $0.16/20 = 0.008$ ). The duration-hazard rate from each 20 minute interval is plotted along with the hazard rate from the fitted Weibull distribution in Figure 8-3. These duration-hazard rates are increasing in both the discrete and continuous formulations. In the discrete calculation, there are fewer data in the right tail of the delay distribution so the 95% confidence interval of the hazard rate increases.

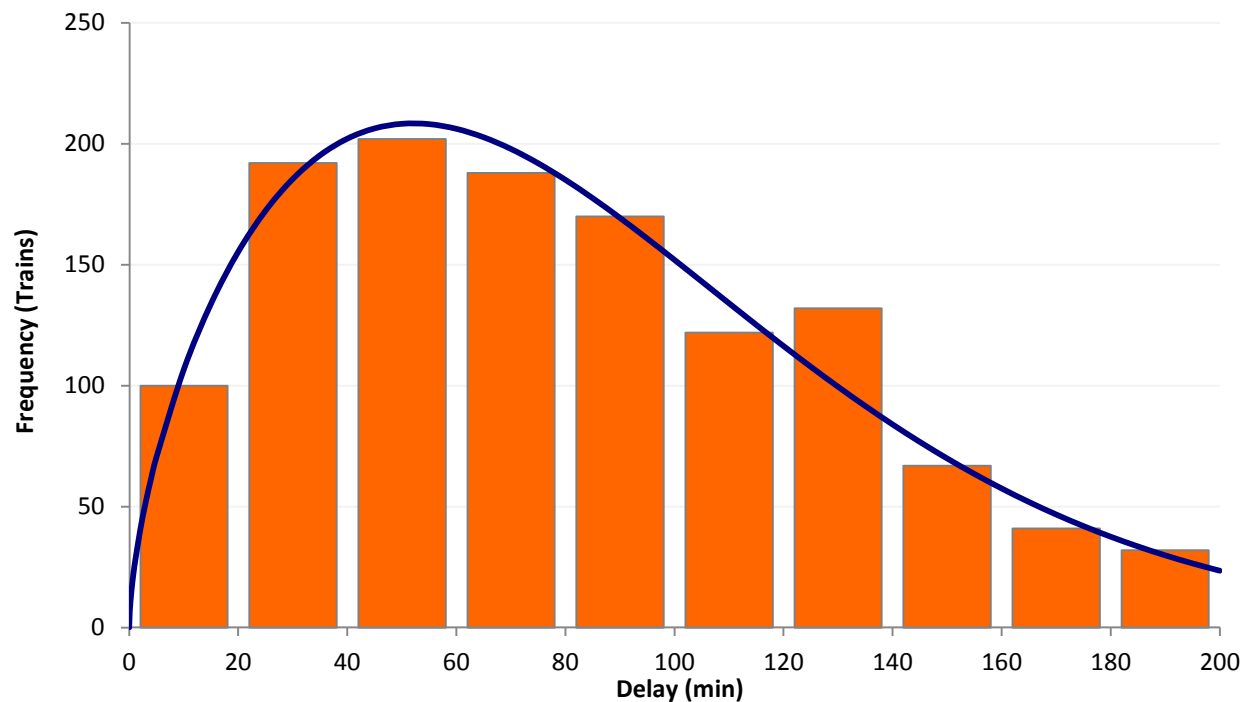


Figure 8-2: Delays following a Weibull distribution (64 TPD, 80% Double track).

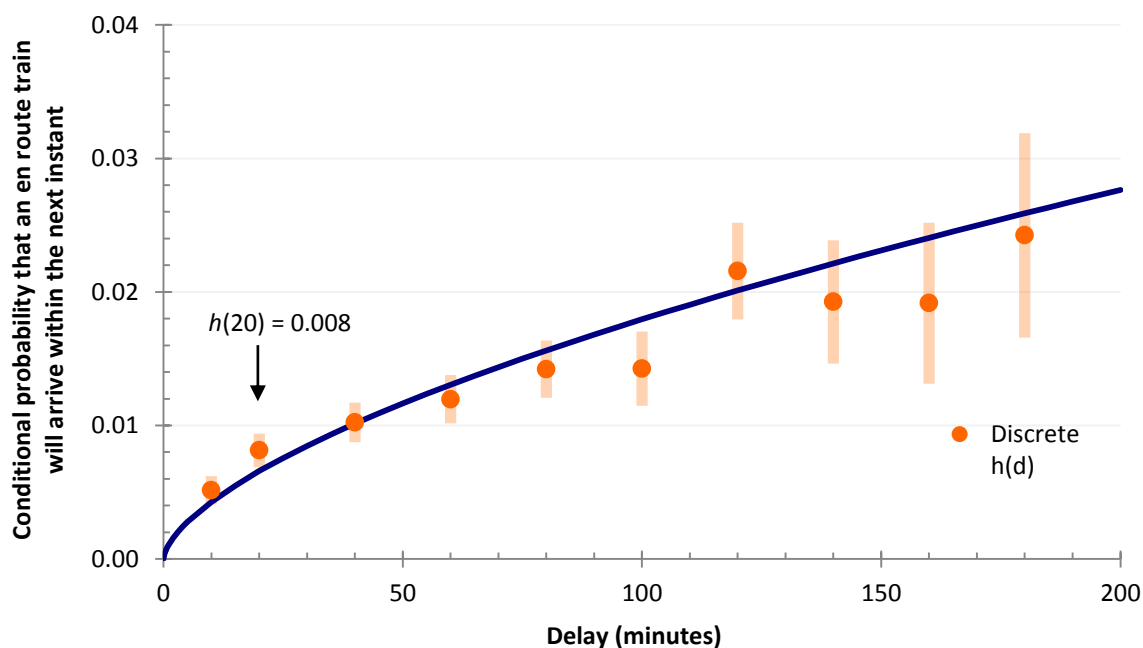
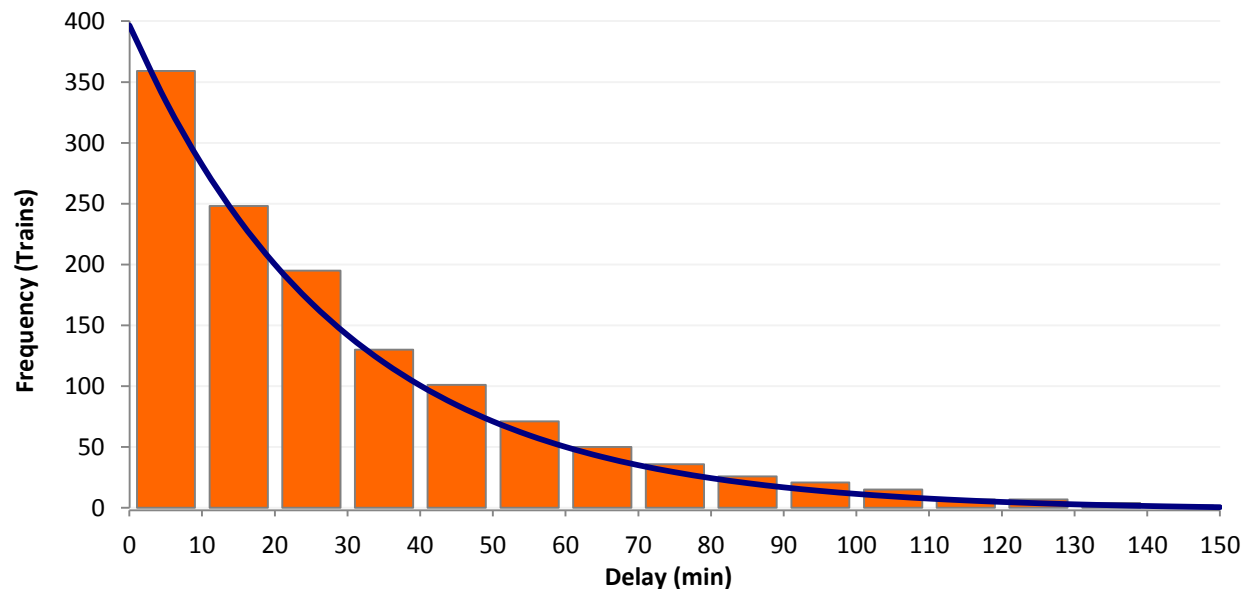


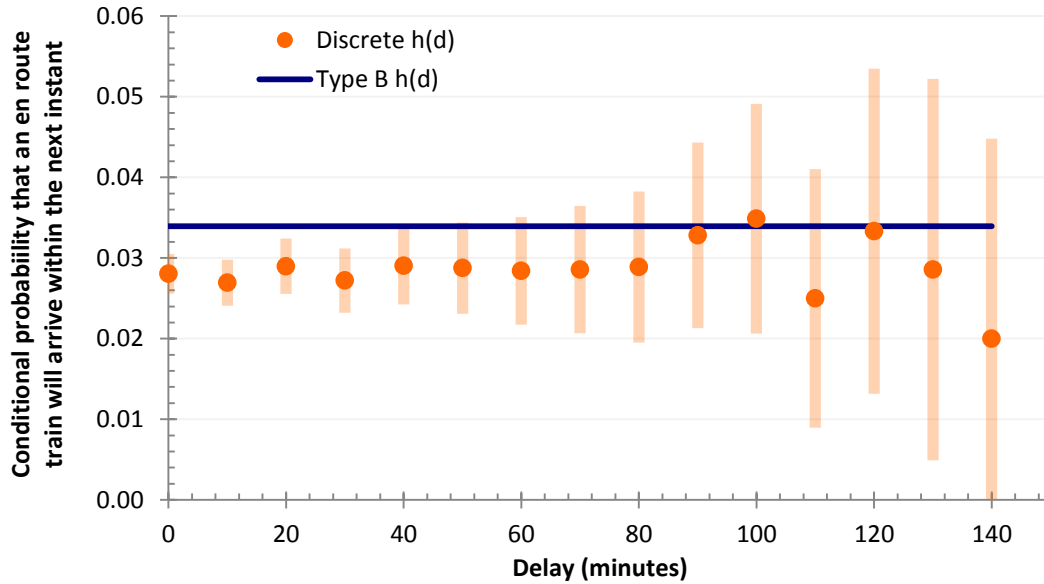
Figure 8-3: Increasing conditional probability that an en route train will arrive within the next minute.

An exponential (“Type B”) distribution fit is shown in Figure 8-4 over a delay population of 1,280 trains from Chapter 7 featuring 64 TPD and 93.2% double track. As before, bars are the

distributions of delay times. The line is the exponential function fit to the data. Most trains perform close to the MRT. For an exponential distribution, the conditional probability that an en route train will arrive within the next minute (duration-hazard rate) is constant. For example, 359 trains were less than 10 minutes late. Of the remaining 921 en route trains, 248 trains will arrive between 10 to 20 minutes late. When an en route train was 10 minutes late, the conditional probability that it will arrive with 10 to 20 minutes of delay was then 0.27 ( $248/921 = 0.27$ ), and the rate per minute was 0.027 ( $0.27/20 = 0.027$ ). The duration-hazard rate for each 10-minute interval is plotted in Figure 8-5, as well as the hazard rate from the exponential distribution. These duration-hazard rates are close to constant in both the discrete and continuous formulations. In the discrete calculation, there are fewer data in the right tail of the delay distribution so the 95% confidence interval of the hazard rate increases.



**Figure 8-4: Train delays following an exponential distribution (64 TPD, 93.2% Double Track).**



**Figure 8-5: Constant conditional probability that an en route train will arrive within the next minute.**

The duration-hazard rate for an exponential (“Type B”) distribution is constant. This distribution has no memory, so what happened in the past has no bearing on the probability of a train arrival in the future. Consider an inbound train that, on average, arrives 30 minutes after its MRT and the delays follow an exponential distribution. Regardless of how late a train is, its probability arriving in the next minute is constant. From the terminal perspective, the lateness of a train does not change the expectations of the yardmaster regarding the likelihood of its arrival.

If the delays follow a Weibull (“Type A”) distribution, very few trains will arrive close to the MRT. Subsequently, the lateness of a train will change the expectations of the yardmaster concerning the likelihood of the train’s arrival. Under this delay distribution, if the MRT of an inbound train corresponded to a 12:00 arrival at a terminal, then at 12:01, the yardmaster has low expectation of the train arriving within the next minute. However, as time goes on and train does not arrive, the conditional probability of the train arriving increases. By contrast to the situation with the exponential distribution (“Type B”), expectation of the train’s arrival is constant.



The duration-hazard rate is most relevant to traffic planners at delay values close to the MRT. When the duration-hazard rate is greater than zero at the MRT, then there is a greater likelihood of trains arriving at the MRT or accruing very little delay. The next important aspect of the duration-hazard-rate function is whether the duration-hazard rate is constant, increasing, or decreasing. This relationship will describe the right tail of the delay distribution. In the case of a constant rate, the same proportion of en route trains will be expected to arrive within the next instant independent of how much delay each train has accumulated. With an increasing hazard rate, then the proportion of en route trains about to arrive is increasing with respect to the remainder of the en route trains. Lastly, with a decreasing rate, the proportion of en route trains about to arrive relative to the remainder of the en route trains is decreasing. A decreasing duration-hazard rate can be associated with “thin” tails of the right side of a delay distribution. This type of relationship may occur on a well-performing double-track line where it is rare for a train to be very late.

An important aspect of reliability analysis is that the cumulative arrival curve and the duration-hazard curve are interrelated (Equation 8-5) (Washington et al. 2003). An analyst can start fitting known probability functions to the distribution or determine a functional relationship from the duration-hazard rate. Once one relationship is assumed, the other is automatically derived. So by assuming a probability distribution, the analyst may be forcing behavior on the hazard function that disagrees with the data. Hazard-based modeling may be advantageous when the trains in the delay distribution have heterogeneous properties. For example, a delay distribution may include both bulk and intermodal trains. Fitting a Weibull distribution to the total delay population ignores information of train type. If the dataset is large enough, it could be split by train type and two unique Weibull distributions could be calculated. However, a hazard-

based approach could model this without splitting the dataset by describing hazard as a function of both *delay*, and *train type* with one equation. Additionally, a duration hazard can model heterogeneous datasets when trains are differentiated by a continuous variable such as horsepower per trailing ton ratio. For example, a delay distribution from manifest trains may include consists with differing tonnage and amounts of motive power. There may not be enough replicates of this continuous variable to split the dataset as was done with the example with different train types. There may have only been 1 train in the dataset that which had a consist with 1.23 horsepower per trailing ton ratio. Duration hazard models could account for different train weights and assigned power in the reliability analysis (Washington et al. 2003, SAS Institute Inc. 2012). For example, a duration-hazard study might find that heavier trains with less power had a greater probability of long delays.

$$F(d) = 1 - \exp \left[ - \int_0^d h(d, X) \right] \quad (8-5)$$

Where:

$F(d)$  = Cumulative arrival function

$h(d, X)$  = Duration-hazard rate function

$d$  = Minutes of delay

$X$  = Vector of train parameters such as type, power, weight, tonnage, origination, and priority

## 8.2 Distributions by Network Configuration

Single-track distributions are likely to follow “Type A” distributions and double-track delay distributions are likely to follow “Type B”. The MRT provides a boundary constraint such that these distributions are asymmetrical and always positive. On a single track line with passing sidings, the “Type A” distribution is derived from the probabilistic nature of a train being the lower priority train and stopping on a siding in a meet conflict with an opposing train. On a homogeneous corridor, a train is expected to be favored in 50% of all meet conflicts; however,

not all delays in meet conflicts are going to be equal. Sometimes the theoretical meeting point between two trains is going to be close to the location of a siding and other times this meet point will be on a single-track section. Additionally, the train that stops on a siding may have to wait for more than one oncoming train. There is usually a minimum delay time to meet an oncoming train, and an unbounded maximum. In a double track configuration, the “Type B” distribution is derived from a portion of trains capable of traveling close to the MRT. Train delays are often caused by traffic travelling against the dominant direction of travel on one of the main tracks. Additionally, faster trains can be delayed by trailing behind slower trains. Small delays are likely and long delays are uncommon (Sogin et al. 2013a).

Connecting pre-existing sidings can increase the probability that the theoretical meeting point between two trains will occur on a section of double track. In this case, neither train accrues delay and it is often referred to as a “rolling meet”. With a percentage of double track installed, some trains may perform close to the MRT and not have any conflicts with trains on the single-track sections. The delays of these trains would follow a “Type B” distribution. The remaining trains would have conflicts on the single-track segments and these train delays would follow a “Type A” distribution. When these two train groups are combined into one dataset, the resulting distribution should still follow a “Type A” or “Type B” distribution. In Section 8.4, the transition from “Type A” to “Type B” is determined as a function of double-track percentage and traffic volume.

### **8.3 Modeling the Delay Distribution Transition**

Similar to increasing capacity by installing sections of double track, the resulting delay distribution from a hybrid-track configuration would be expected to switch from “Type A” to “Type B”

as the amount of double track is increased. The dataset from Chapter 7 was used in the subsequent analysis. This dataset includes the delays of freight trains operating in 200 different simulations representing various permutations of the factors describing the amount of double track, trains per day, and the type of interfering traffic (Table 8-1). Using JMP (2013), both “Type A” and “Type B” distributions were fitted to the delay data from each simulated permutation in Table 8-1. The quality of the fit for each distribution was determined visually through diagnostic plots, fit statistics, and the position of the lower quantiles. When the minimum, or 10% quantile was close to zero, then the distribution was more likely to be “Type B”. Additionally, a “Type B” distribution may be likely when the shape parameter of the fitted Weibull distribution is close to 1.0. The Weibull distribution is a robust distribution capable of fitting a wide range of datasets when compared to the exponential distribution. However, an exponential distribution may be more parsimonious in certain cases because it is a simpler model with fewer parameters.

**Table 8-1: Factors That May Alter Delay Distributions.**

Factor	Range
Percentage of Double Track	19% - 100%
Traffic Volume (TPD)	8 – 72
Type of interfering traffic	100% Freight (homogenous) 75% Freight, 25% Passenger limited to 50 mph 75% Freight, 25% Passenger limited to 110 mph

Logistic regression is commonly used to predict a binary response from an array of inputs. The logistic regression procedure is in the form of Equation 8-6a, which calculates the probability of a positive response. For this analysis, a negative response was defined as the case when the delay distribution followed a “Type A” distribution and a positive response was defined as the case when the distribution follows a “Type B” distribution with a portion of trains

performing at or close to the MRT. Equation 8-7 was the best model using the logistic regression procedure from JMP. The parameter statistics are summarized in Table 8-2. The type of interfering traffic was not statistically significant and not included in the model.

$$\pi = \frac{1}{1 + e^{-f(X)}} \quad (8-6a)$$

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots \quad (8-6b)$$

$$f(X) = -15.36 + 38.33(D.T. \%) - 0.4096(Vol) + 0.8422(D.T. \%)(Vol) \quad (8-7)$$

Where:

$\pi$  = Probability of delays following a “Type B” distribution

(D.T. %) = Double-track percentage

(Vol) = Volume

**Table 8-2: Parameters of Logistic Regression.**

Term	Estimate	ChiSquare	Prob>ChiSq	Lower 95%	Upper 95%
Intercept	-15.36	22.21	<.0001*	-23.05	-10.02
Double Track %	38.33	23.87	<.0001*	25.65	57.07
Volume	-0.4096	22.12	<.0001*	-0.62	-0.27
(Double Track %)*(Volume)	0.8442	18.25	<.0001*	0.51	1.30

Once a model predicts a probability,  $\pi$ , then a probability threshold needs to be established in order to classify the data. Different thresholds can lead to different false-positive and false-negative rates. Receiver-operating-curve (ROC) analysis can evaluate the quality of a model by analyzing many threshold values simultaneously. The area under the ROC curve of this model was 0.992 indicating a good fit to the data. The highest percent-correct rating occurred when the probability threshold was 0.526, with probabilities greater than this value indicating that the delays followed a “Type B” distribution. The accuracy of the model had low sensitivity to the selection of the probability thresholds. When potential threshold values were between 0.08 and 0.93, the models would have at least 90% of the cases identified correctly.

With greater double-track percentage, and lower traffic volume, a delay distribution will more likely follow an exponential distribution than a “Type B” distribution (Figure 8-6). Similar to the models developed in Chapter 6 and Chapter 7 there was a strong interaction between double-track percentage and traffic volume. In each of the candidate models, the type of interfering traffic was always insignificant. The curve plotted in Figure 8-6 shows the transition zone from “Type A” (Weibull) to the “Type B” (exponential) distribution. Out of the two-hundred distributions, only seven were misclassified by the logistic regression model. These seven cases fell on the transition curve in Figure 8-6.

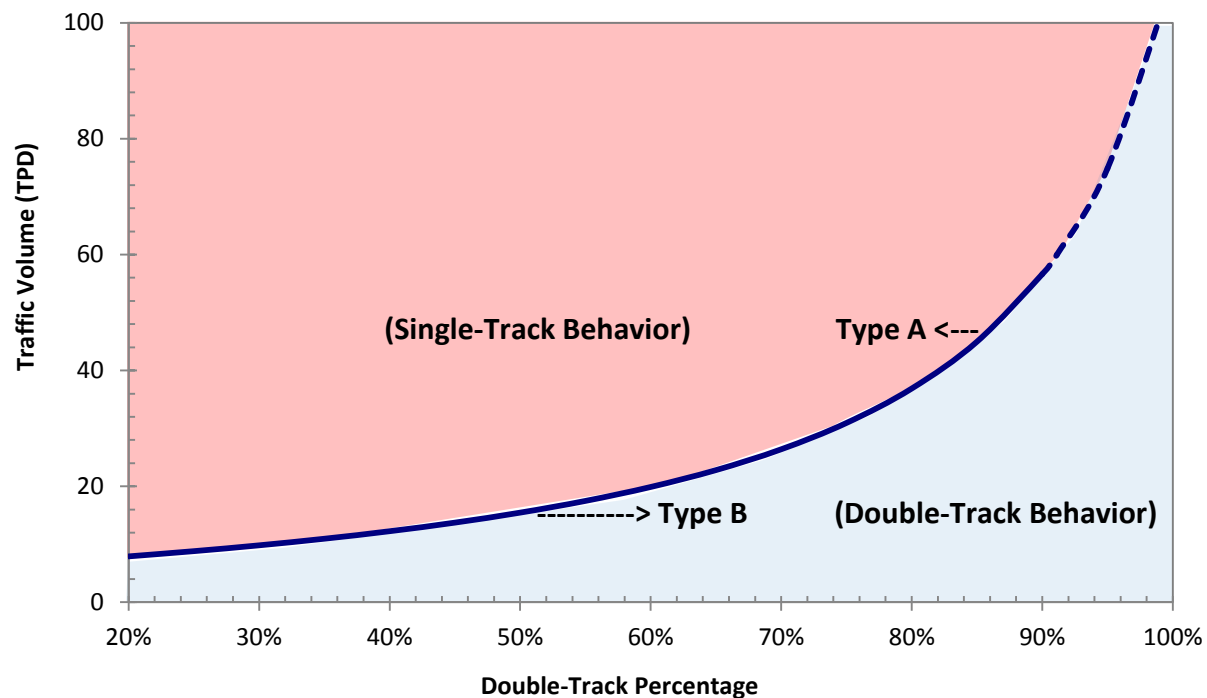


Figure 8-6: Estimated transition function from delay behavior similar to single track and behavior similar to double track (Derived from Equation 8-6a and 8-7 where  $\pi = 0.53$ ).

## 8.4 General Findings & Future Work

The interaction between double track and traffic volume in this chapter is similar to the trends found in Chapter 6 using design of experiments techniques, and Chapter 7 using non-linear

regression. These all indicate that the benefits of double-track start to compound after 60% of the route is double track. This model was developed on freight dominated corridors with 100% freight trains and 75% freight trains. This small change in heterogeneity may not have been large enough to capture a change in delay distributions. The type of distribution may be useful in determining practical capacity utilization rates for railway infrastructure as well as setting reasonable on-time performance metrics. Duration-hazard rates may be useful for tactical planning of terminals with regard to variation in train arrivals. While the Weibull distribution is robust for fitting delay distributions from homogenous trains, a duration-hazard model may help model the effects of heterogeneity. Additionally, the presented concepts of failure (train arrival) rates may supplement models for railcars making connections at terminals and classification yards.

## **CHAPTER 9**

### **GENERAL FINDINGS AND FUTURE RESEARCH OPPORTUNITIES**

#### **9.1 Future Research**

The research methodologies and analysis techniques presented in this thesis can be extended to look at a more comprehensive list of factors or study new railway capacity questions. The following section highlights future areas of research.

##### **9.1.1 Single-Track**

Bottleneck analyses for single-track generally identify the pair of adjacent sidings with the longest transit time between them. However, there may be other factors affecting the location of the bottleneck besides longest gap between sidings. Focusing on just one pair of sidings, may cause other global characteristics to be missed. Chapter 3 describes a zone analysis approach to identification of bottlenecks. A systematic series of simulation experiments could analyze various combinations of typical railroad configurations and identify the bottlenecks through this type of zone analysis. Some bottleneck locations may not correspond to the longest travel time between sidings.

An important consideration for freight railroads is train length. Longer trains can reduce the number of trains on a corridor and improve the efficiency of locomotive, crew, and fuel use. However, the sidings on a single-track railroad may not be long enough to accommodate long trains. Running a small percentage of the trains longer than the siding size may delay the need for capital expenditure on the corridor. This strategy may be feasible if the long train does not conflict with another long train on a section of single-track. Subsequently, long sidings may be needed to resolve conflicts between long trains. A simulation experiment would develop curves



relating daily railcar throughput to the long-train percentage of total traffic and the long-siding percentage of total number of sidings. An important consideration is that these long trains would require priority to hold the main at meet points so dispatching would have to be carefully planned to avoid exacerbating delays.

### **9.1.2 Partial Double Tracking**

Equation 7-10 is a useful tool for analyzing the change in capacity by adding sections of double track. Only three different traffic scenarios were compared: homogenous freight, and two mixed freight and passenger scenarios. More traffic scenarios could be investigated using design of experiments technique. Each run in the experimental matrix would correspond to a suite of simulations that would produce a unique form of Equation 7-10. A minimum of three data points are needed to estimate the three-parameter estimates of Equation 7-10; however, more points may be needed to achieve stable estimates. From this framework, more factors could be investigated and related back to delay-volume curve analyses and the benefits of partial-double tracking.

More investigation should be conducted into the validity of Equation 7-9. There may be other scenarios in which the linear relationship between double-track percentage and delay does not hold. This relationship may be disproved by relaxing the constant 10-mile siding spacing assumption. Repeating the experiments under a longer siding spacing may lead to a different slope-term of delay reduction per mile of double track (Equation 9-1). Connecting a larger gap will increase the denominator,  $x$ , of the slope equation. However, there might be greater delay reductions by removing a long bottleneck section and consequently increasing the numerator of the slope equation. While these two factors counteract each other, the working hypothesis is that that there would be disproportionately larger delay reductions by double tracking over a greater

siding spacing increment. If this hypothesis is correct, then systematically connecting the single-track sections in order of longest to shortest may yield a non-linear curve. This curve would be expected to be concave with greater delay reductions at lower double-track-percentages (Figure 9-1).

$$\gamma_1(V) = \frac{\Delta D}{\Delta x} \quad (9-1)$$

Where:

$\gamma_1(V)$  = Reduction in train delay per double-track percentage point (or mile)

$D$  = Train Delay

$x$  = Double-track percentage (or miles of double track)

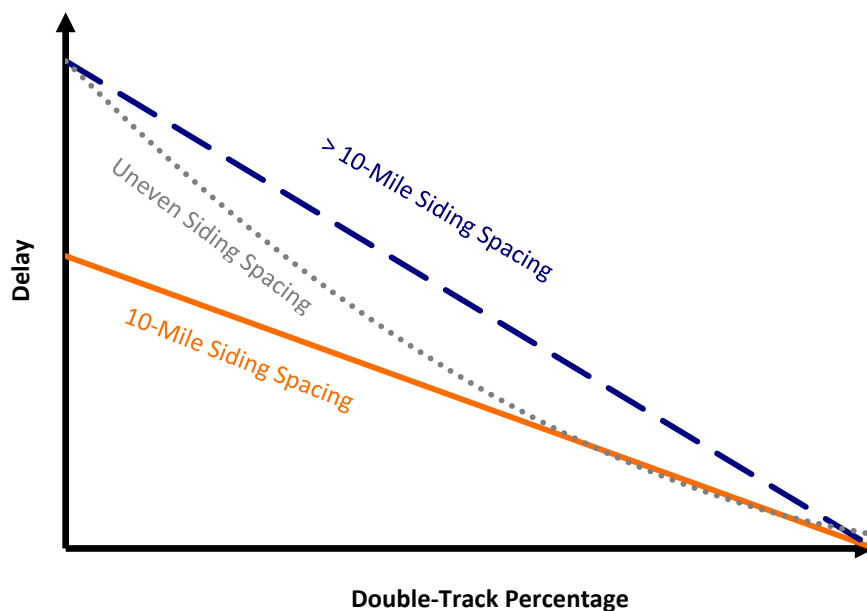


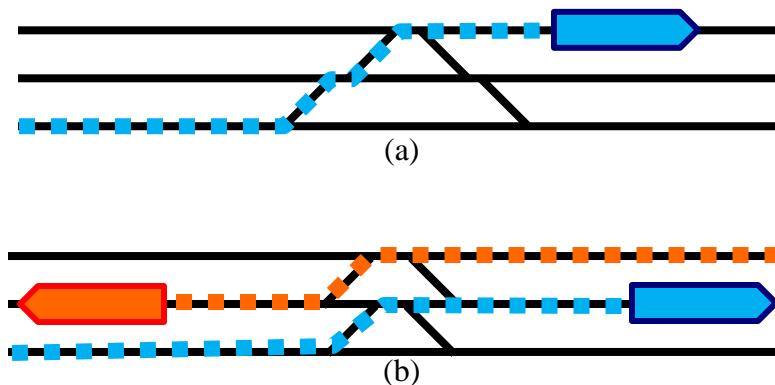
Figure 9-1: Hypothesis on non-linear delay reductions from double tracking under constant volume.

### 9.1.3 Triple Track

The capacity relationships developed in Chapters 6 through 8 focused on the incremental transition from single to double track line by systematically connecting pre-existing passing sidings. These methodologies could be applied towards the incremental transition from double to triple track. Faster trains can overtake slower trains without causing delays to trains traveling in the opposite direction. If the third track is long enough, the faster train can overtake the slower

one without the slower train stopping. Similar to double tracking, the capital for the full triple-track configuration may not be available at once. Having overtaking sidings or partial sections of triple track may be intermediate solutions.

An interesting aspect to triple track is the configuration of crossovers. In general, the crossovers can be *universal* where a train traveling in either direction can transfer to either of the tracks. Alternatively, the crossovers can be *parallel* where multiple trains can use the interlocking simultaneously. Figure 9-2a illustrates a universal crossover where an eastbound train on the bottom track shifts to the top track and blocks all other movement through the interlocking for other trains. In Figure 9-2b, the crossovers are *parallel* where two trains can shift over one track simultaneously. An intensive simulation study may indicate which crossover strategy is appropriate under different scenarios. *Universal* configurations may be more appropriate when trains need to be sorted entering or leaving a terminal area whereas a *parallel* configuration may have greater efficiencies with high speed mainline movements between cities.



**Figure 9-2: Crossovers in (a) *universal* and (b) *parallel* configurations.**

The location of existing CTC control points and universal crossovers may influence future infrastructure configurations. A potential low cost solution could be installing triple track between two universal crossovers and use the existing signaling infrastructure. In the case of

installing overtaking sidings, it is unlikely that the overtaking sidings would be long enough to span the distance between adjacent control points. Only one end of the overtaking sidings may be placed at an existing control point of a universal crossover while the other end of the siding would need a new control point. Figure 9-3 depicts five different arrangements for constructing an overtaking siding in which the left end is at the existing control point. Configurations A and B limit route options by not having crossovers on the east end of the siding between the top and bottom tracks. Configurations A and C allow *parallel* movements while Configurations B and D feature *universal* movements. Configuration E has the overtaking siding installed between the existing main tracks such that trains can enter the siding from either direction without conflicting with the other main track. Exploratory simulation experiments can give insight on the capacity benefits of each configuration.

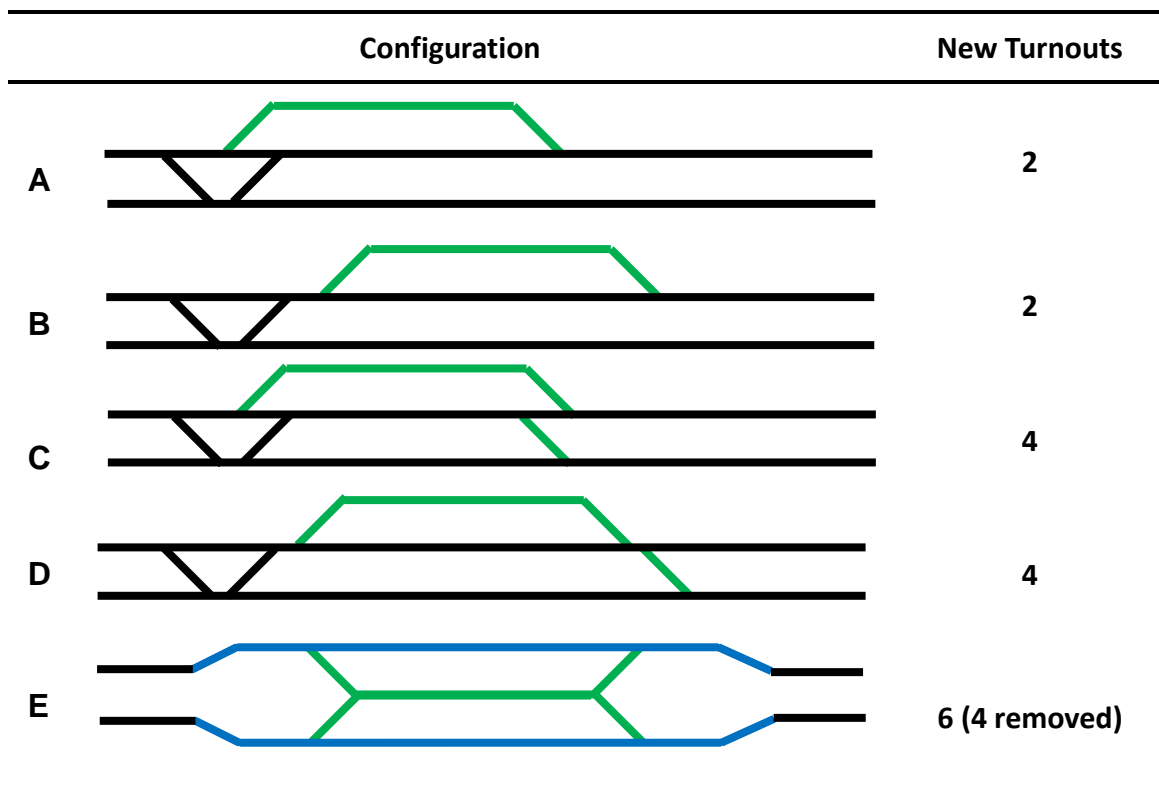


Figure 9-3: Overtaking Siding Configurations.

The triple track could be located along the line using the same allocation methods described in Chapter 6, *Alternate*, *Split*, and *Grouped*. These triple tracking strategies can be investigated using methodologies developed in Chapter 6 using design of experiments or non-linear regression similar to the methodology used in Chapter 7. Additionally, many sections of triple track are associated with commuter rail lines close to city centers. Interesting commuter-rail specific factors may include stopping patterns, stop spacing, peak and off peak intensity.

#### **9.1.4 Heterogeneity in Train Types**

Most of the work presented in this thesis involved higher-speed passenger trains and freight trains sharing infrastructure. The methodologies described could be adapted to analyze interactions between other train types such as bulk and intermodal or commuter and freight trains. The traffic density of commuter trains during morning and evening rush-hours can prevent freight trains from using the line. Freight traffic is often held in holding yards until the rush period is over. When it finishes, the freight trains can use the slots in-between the off-peak frequencies of the commuter trains. Some factors to investigate include the size, and location of the holding yard, as well as the rush-hour frequency, off-peak frequency, maximum speeds, infrastructure layout, and station-stop spacing. An important output would be how many freight trains can depart a holding yard in-between the off-peak frequencies.

Another source of heterogeneous interactions is that of full and empty coal trains on a steep grade. There are many mine-to-port coal operations where coal is mined at higher elevations and transported downgrade to a harbor for export. There may be an optimal assignment of power and priority to the full and empty trains that maximizes the fluidity of the operation. Variables to be investigated may include, grade, power, tonnage, maximum speed,

priority, and single track versus double track. This experiment could also be repeated to incorporate steep positive grades in both directions.

### **9.1.5 Analysis Methods**

Design of experiments (DOE) techniques are powerful tools to analyze numerous factors that may affect railway capacity. In Chapter 6, five factors influencing a “hybrid” track configuration were studied in a partial factorial experimental design. An extension for DOE work is to investigate more factors that may affect capacity such as grades, curvature, and parameters of the train types such as power, weight, and length. By doing a screening analysis of 10 to 20 factors, there may be surprising interactions between variables and new insights gained. Alternatively, a DOE approach would be interesting in the context of project selection. Suppose there is a series of five different infrastructure upgrades along a corridor but not all may be needed to accommodate future traffic. If three traffic scenarios considered were current traffic, low-growth, and high-growth then there would be 96 ( $2^5 \times 3$ ) runs for the full-factorial experiment. A partial factorial of 30 simulations may also be feasible for the analysis. The final regression model would be a dynamic communication tool for various stakeholders. They could dynamically change traffic volume, select projects, and see the corresponding delay changes for each train type. This type of project selection framework may improve the efficiencies of simulation studies.

Another area of future research is applying ordinal, logistic modeling to simulation results. This would gain more information from scenarios that do not have feasible dispatching solutions. Many high traffic and low trackage cases will be infeasible for the simulation software to derive a dispatching solution. These infeasible cases are not used in the final data analyses; however, there is still information to be gained from the fact that they were infeasible. Ordinal

logistic regression could classify results as “low”, “medium”, “high”, and “infeasible”. Despite losing resolution of the response surface, this method leverages more information from the data output of the simulation software.

#### **9.1.6 Capacity Project Selection for Passenger Services**

Modeling passenger ridership is difficult; knowing how many people will use a rail service before it is operational is non-trivial. There are many non-tangible factors such as accessibility, demographics, station connectivity, and current travel patterns. However, the core of a ridership model includes four key metrics: travel time, frequency, reliability and cost. In many cases, the fare will be derived from the user costs of competing transportation modes. The remaining three metrics are interrelated by the railway operation. Passengers are expected to prefer high frequencies to maximize flexibility as well as low travel time, and high reliability. The capacity of the corridor is an important factor governing frequency. Adding more trains without investing in capacity may increase congestion, decrease reliability, and decrease travel time (AREA 1931). Increasing average speeds to reduce run time, may increase delays to freight trains, and may warrant capacity upgrades to mitigate this additional delay. Corridor improvement projects address these three metrics differently. A fundamental understanding of these operational tradeoffs is needed to supplement the ridership model such that projects can be prioritized.

Project selection models can be used to prioritize projects under various constraints. A future corridor investment model would maximize the “Transportation Utility Engine”. There are inherent tradeoffs between faster trains, more frequency, and greater reliability. Caughron (2013) developed a project selection optimization model to enable selection of the most cost-effective infrastructure upgrades to meet various minimum-run-time targets. In his case studies, cost was an exponential function of travel time reduction. In my work, capacity could be measured by

Equation 7-10. These two models could potentially form the framework of a future optimization model that balances the tradeoff between greater frequency by partial double tracking versus the alternative of faster travel times by track structure and alignment improvements (Figure 9-4). The objective of such a model could be to maximize ridership or maximize net-present-value.

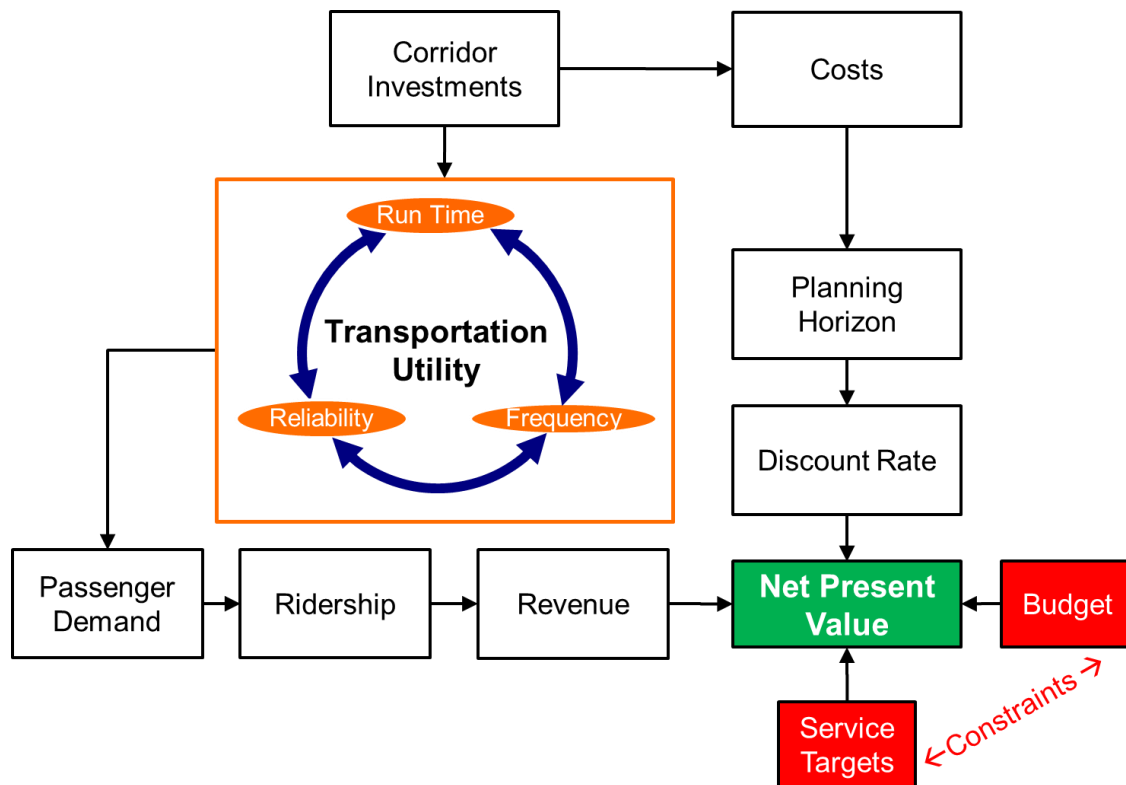


Figure 9-4: Shared Corridor Optimization Model (Caughron 2013).

## 9.2 Conclusion

Capacity can be defined in several ways including increased throughput, reliability, and asset utilization. It can be a somewhat subjective metric that is analyzed using various techniques. The highway mode has developed key traffic relationships that can be adapted to railroad traffic. Railroads share characteristics that are similar to fundamental highway traffic relationships, but also have unique traffic relationships that are specific to railway operating patterns, infrastructure



and control systems. A standardized set of railroad capacity and performance metrics may offer considerable benefits to rail traffic planners helping them more effectively communicate the ability of the rail network to support future traffic growth.

Sharing tracks with higher speed intercity passenger and freight trains is a challenge. In all cases studied, the addition of passenger trains increased freight train delay but the mechanisms differed between single and double track configurations, freight and passenger dominated lines, and passenger train speed. On single-track lines, the greatest impact to freight trains was due to passenger trains having higher priority in meets, rather than their speed differential. By contrast, a double-track configuration nearly eliminated meet delay, but the higher speed differential between train types caused more overtake conflicts. Increasing passenger train speed reduces travel time, but may also decrease reliability.

Partial-factorial and response surface modeling can be used to quickly compare relative effects in railway simulation analyses. If a single-track corridor is being upgraded to a double-track corridor, major factors influencing capacity include the amount of double track installed, and the minimum level of service (MLOS). The distribution of double track along the route can have a modest effect on train delays. Minor effects included the passenger train speed and the heterogeneity of the line. The capacity transition from single to double was approximated by a convex (concave upward) function in which the last miles installed have a much greater effect on increasing capacity.

Complementing the work from Chapter 6, non-linear regression over a higher resolution data set found that train delays decrease linearly with additional sections of double track when the volume is constant. Under a constant MLOS, the incremental capacity gained from each additional section of double track increased. The simultaneous operation of freight and passenger

trains on the same trackage reduced the corridor's capacity as well as the incremental capacity benefits from installing segments of double track. The marginal loss in capacity from heterogeneous operation was greater on lines close to full double track than on hybrid track configurations that were closer to single-track lines. When large speed differentials were present between train types, the speed differential may not be a significant delay causing mechanism until most of the line was double tracked.

These analyses all indicate that double-tracking benefits start to compound after 60% of a line is double track where it is more likely for delays to follow an exponential distribution. Understanding this distribution may be useful in determining practical capacity utilization rates for railway infrastructure, as well as setting reasonable on-time performance metrics. Concepts from failure analysis can be applied to railroad train delays to describe reliability of the transportation service provided.

## REFERENCES CITED

- AAR. 2013. *U.S. Rail Traffic: Week 12, 2013*. Association of American Railroads, Washington, D.C.
- AASHTO. 2007. *America's Freight Challenge*, System. American Association of State Highway and Transportation Officials, Washington, D.C.
- Abril, M., F. Barber, L. Ingolotti & M. Salido. 2008. An Assessment of Railway Capacity. *Transportation Research Part E: Logistics and Transportation Review* 44: 774–806.
- Al-Deek, H. & E. Emam. 2006. New Methodology for Estimating Reliability in Transportation Networks with Degraded Link Capacities. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 10: 117–129.
- AREA. 1921. Notes on the Determination of the Traffic Capacity of Single and Multiple Track Railways. In: *Proceedings of the 1921 Convention of the American Railway Engineering Association*. Chicago, IL, pp. 744–759.
- AREA. 1931. Method of Increasing the Traffic Capacity of a Railway. In: *Proceedings of the 1931 Convention of the American Railway Engineering Association*. Chicago, IL, pp. 643–692.
- Assad, A.A. 1980. Models for Rail Transportation. *Transportation Research Part A: General* 14: 205–220.
- Benekohal, R.F. & W. Zhao. 2000. Delay-Based Passenger Car Equivalents for Trucks at Signalized Intersections. *Transportation Research Part A: Policy and Practice* 34: 437–457.
- Bhat, C.R. 1996. A Hazard-Based Duration Model of Shopping Activity with Nonparametric Baseline Specification and Nonparametric Control for Unobserved Heterogeneity. *Transportation Research Part B: Methodological* 30: 189–207.
- Biehler, A.D., L.L. Brown, W.A. V. Clark, L. Angeles & D.S. Ekern. 2008. *Cost-Effective Performance Measures for Travel Time Delay, Variation, and Reliability*, Planning And Administration. National Cooperative Highway Research Program, Washington, D.C.
- Bing, A.J., E.W. Beshers, M. Chazvez & D.P. Simpson. 2010. *Guidebook for Implementing Passenger Rail Service on Shared Passenger and Freight Corridors*. National Cooperative Highway Research Program, Washington D.C.
- Burdett, R.L. & E. Kozan. 2006. Techniques for Absolute Capacity Determination in Railways. *Transportation Research Part B: Methodological* 40: 616–632.
- Cambridge Systematics. 2007. *National Rail Freight Infrastructure Capacity and Investment Study*. Cambridge Systematics, Washington, D.C.

- Carey, M. 1994. Stochastic Approximation to the Effects of Headways on Knock-On Delays of Trains. *Transportation Research Part B: Methodological* 28: 251–267.
- Caughron, B.M. 2013. Rail Corridor Project Selection Model. In: *Proceedings of the 2013 Joint Rail Conference*. ASME, Knoxville, TN. (Presentation).
- Chen, B. & P.T. Harker. 1990. Two Moments Estimation of the Rail Lines on Single-Track with Scheduled. *Transportation Science* 24: 261–275.
- Dingler, M.H., Y.-C. Lai & C.P.L. Barkan. 2009. Impact of Train Type Heterogeneity on Single-Track Railway Capacity. *Transportation Research Record: Journal of the Transportation Research Board* 2117: 41–49.
- Dingler, M.H., A. Koenig, S.L. Sogin & C.P.L. Barkan. 2010. Determining the Causes of Train Delay. In: *Proceedings of the Annual AREMA Conference*. Orlando, FL.
- Dingler, M.H., Y.-C. Lai & C.P. Barkan. 2013a. Effect of Train-Type Heterogeneity on Single-Track Heavy Haul Railway Line Capacity. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 10.1177/09.
- Evans, M., N. Hastings & B. Peacock. 2000. *Statistical Distributions*. John Wiley & Sons, Inc., New York.
- Federal Railroad Administration. 2013a. Obligated High Speed Intercity Passenger Rail Funding by Region [WWW Document]. URL <http://www.fra.dot.gov/Page/P0554>
- Federal Railroad Administration. 2013b. Title 49 - Part 228: Hour of Service of Railroad Employees. United States of America Code of Federal Regulations, Washington D.C.
- Federal Railroad Administration. 2013c. *Title 49 - Part 236 Part D: Traffic Control Systems*. United States of America Code of Federal Regulations, Washington, D.C., United States of America Code of Federal Regulations.
- Galaverna, M. & G. Sciutto. 1999. Influence Traffic Composition on the Capacity of mixed-traffic Railway Lines. *Influence of Traffic Composition on the Capacity of Mixed-traffic Railway Lines* 28: 14–16.
- Galloway, D. 2002. Train Performance Characteristics. In: *Transportation Research Board Workshop on Railroad Capacity and Corridor Planning*. Transportation Research Board (TRB), Washington, D.C.
- Goos, P. & B. Jones. 2011. *Optimal Design of Experiments: A Case Study Approach*. John Wiley & Sons, Ltd., West Sussex.
- Greenshield, B.D. 1935. A Study of Traffic Capacity. In: *Proceedings of the 14th Annual Meeting of the Highway Research Board*. Washington, D.C., pp. 448–477.

- Hallowell, S.F. & P.T. Harker. 1998. Predicting On-Time Performance in Scheduled Railroad Operations: Methodology and Application to Train Scheduling. *Transportation Research Part A: Policy and Practice* 32: 279–295.
- Harrod, S. 2009. Capacity Factors of a Mixed Speed Railway Network. *Transportation Research Part E* 45: 830–841.
- Higgins, A., E. Kozan & L. Ferreira. 1996. Optimal Scheduling of Trains on a Single Line Track. *Transportation Research Part B: Methodological* 30: 147–161.
- Huisman, T. & R.J. Boucherie. 2001. Running Times on Railway Sections with Heterogeneous Train Traffic. *Transportation Research Part B: Methodological* 35: 271–292.
- Ireland, P., R. Case, J. Fallis & J. Kuehn. 2003. Perfecting the Scheduled Railway: Model-Driven Operating Plan Development. In: *System*. Phoenix, AZ, pp. 1–28.
- Jelaska, M. 2000. An Interaction Model for Passenger and Freight Traffic on a Single-Track Railway. In: *Proceedings of the 7th International Conference on Computers in Railways*. Bologna, Italy.
- JMP. 2013. *JMP 10*. SAS Institute INC., Cary, NC.
- Jones, B., L. Janssen & F. Mannering. 1991. Analysis of the Frequency and Duration of Freeway Accidents in Seattle. *Accident Analysis and Prevention* 23: 239–55.
- Kahn, A.M. 1979. *Railway Capacity Analysis and Related Methodology*. Canadian Transport Commission, Ottawa, Canada.
- Kiefer, N.M. 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature* 26: 646–679.
- Kraft, E.R. 1982. Jam Capacity of Single Track Rail Lines. *Transportation Research Forum* 23: 153–162.
- Krueger, H. 1999. Parametric Modeling in Rail Capacity Planning. In: *Proceedings of the 1999 Winter Simulation Conference*. IEEE, Phoenix, AZ, pp. 1194–1200.
- Kulick, B.C. 2011. Operations Analysis Using Simulation Modeling to Evaluate Capacity Requirements for Direct Intermodal Rail Facilities at the Port of Long Beach. In: *Proceedings of the Annual AREMA Conference*. Minneapolis, MN.
- Kutner, M.H., C.J. Nachtsheim, J. Neter & W. Li. 2004. *Applied Linear Statistical Models*, 5th ed. Irwin.
- Kwoka, J. & L. White. 1997. Manifest Destiny? The Union Pacific-Southern Pacific Merger. In: J.E. Kwoka & L.J. White (Eds.), *The Antitrust Revolution: Economics, Competition, and Policy*. Oxford University Press, New York, NY, pp. 27–51.

- Lai, Y.-C. & C.P.L. Barkan. 2011. Comprehensive Decision Support Framework for Strategic Railway Capacity Planning. *Journal of Transportation Engineering* 137: 738–749.
- Lai, Y.-C. & P. Huang. 2012a. High Speed Route Improvement Optimizer. In: *Proceedings of the 91st Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Lai, Y.-C. & Y.-A. Huang. 2012b. Estimation of Single and Double-Track Capacity with Parametric Models. In: *Proceedings of the 91st Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Lai, Y.-C., Y.-H. Liu & T.-Y. Lin. 2012. Development of Base Train Equivalents to Standardize Trains for Capacity Analysis. *Transportation Research Record: Journal of the Transportation Research Board* 2289: 119–125.
- Lai, Y.-C. & M.-C. Shih. 2010. Strategic Capacity Planning Process for Intercity Passenger Rail. In: *Proceedings of the 2010 Joint Rail Conference*. ASME, Urbana, IL.
- Leeson, F. 2002. Old Rails Could Carry Commuters in Future. *The Oregonian* C01.
- Leilich, R.H. 1998. Application of Simulation Models in Capacity Constrained Rail Corridors. In: *Proceedings of the 1998 Winter Simulation Conference*. IEEE, Orlando, FL, pp. 1125–1133.
- Lewellen, M. & K. Tumay. 1998. Network Simulation of a Major Railroad. In: *Proceedings of the 1998 Winter Simulation Conference*. Washington, D.C., pp. 1135–1138.
- Lindfeldt, A. 2012a. *Congested Railways: Influence of Infrastructure and Timetable Properties on Delay Propagation*. Royal Institute of Technology, Department of Transport, Stockholm, Sweeden. (Master's Thesis).
- Lindfeldt, O. 2010. *Railway Operation Analysis: Evaluation of Quality, Infrastructure and Timetable on Single and Double-Track Lines with Analytical Models and Simulation*. Royal Institute of Technology, Department of Transport, Stockholm, Sweeden. (PhD Disertation).
- Lindfeldt, O. 2012b. From Single to Double Track: Effects of Alternative Extension Measures. In: *13th International Conference on Design and Operation in Railway Engineering*. New Fores, Hampshire, U.K., pp. 313–324.
- Lu, Q., M. Dessouky & R.C. Leachman. 2004. Modeling Train Movements Through Complex Rail Networks. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 14: 48–75.
- Martland, C.D. 2008. Improving on-time performance for long-distance passenger trains operating on freight routes. *Journal of the Transportation Research Forum* 47: 63–80.

- Mattsson, L.-G. 2007. Railway Capacity and Train Delay Relationships. In: A.T. Murray & T.H. Brubescic (Eds.), *Critical Infrastructure: Advances in Spatial Science*. Springer, Berlin, Germany, pp. 129–150.
- McClellan, J. 2006. Railroad Capacity Issues. In: *Proceedings of the 84th Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Mitra, S., D. Tolliver & S. Mitra. 2010. Estimation of Railroad Capacity Using Parametric Methods. *Journal of the Transportation Research Forum* 49: 111–126.
- Murali, P., M. Dessouky, F. Ordóñez & K. Palmer. 2010. A Delay Estimation Technique for Single and Double-Track Railroads. *Transportation Research Part E: Logistics and Transportation Review* 46: 483–495.
- O'Connor, P. & A. Kleyner. 2011. *Practical Reliability Engineering, 5th ed.* Wiley, New York, NY.
- Oh, K.K., C.D. Martland & J.M. Sussman. 1995. Origin-to-Destination Trip Times and Reliability of Rail Freight Service in North American Railroads. *Transportation Research Record: Journal of the Transportation Research Board* 1489.
- Pachl, J. & T. White. 2004. Analytical Capacity Management with Blocking Times. In: *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Pachl, J. 2009. *Railway Operation and Control, 2nd ed.* VTD Rail Publishing, Mountlake Terrace, WA.
- Pawar, S.P.S. 2011. *An Analysis of Single Track High Speed Rail Operation*. University of Birmingham, Department of Railway Systems Engineering and Integration, Birmingham, U.K. (Master's Thesis).
- Petersen, E.R. 1982. A Structured Model for Rail Line Simulation. *Transportation Science* 8: 192–206.
- Petersen, E.R. & A.J. Taylor. 1987. Design of Single-Track Rail Line for High-Speed Trains. *Transportation Research Part A: General* 21.
- Poole, E. 1962. *Costs - A Tool for Railroad Management*. Simmons-Boardman Pub. Corporation, New York, NY.
- Pouryousef, H. & T. White. 2013. Review of Capacity Measurement Methodologies; Similarities and Differences in the U.S. and European Railroads. In: *Proceedings of the 92nd Annual Meeting of The Transportation Research Board*. Washington, D.C., pp. 1–14.
- Prokopy, J.C. & R.B. Rubin. 1975. Parametric Analysis of Railway Line Capacity. DOT-FRA-OPPD-75-1. Federal Railroad Administration, Washington, D.C.

- Saat, M.R. & C.P.L. Barkan. 2013. *Investigating Technical Challenges and Research Needs Related to Shared Corridors for High-Speed Passenger and Railroad Freight Operations*. DOT/FRA/ORD-13/29. Federal Railroad Administration, Washington, D.C.
- Sall, J., A. Lehman, M. Stephens & L. Creighton. 2012. *JMP Start Statistics: A Guide to Statistics and Data analysis Using JMP, Fifth Edit ed*. SAS Institute Inc., Cary, NC.
- SAS Institute Inc. 2012. *JMP 10: Quality and Reliability Models*. SAS Institute Inc., Cary, NC.
- Schlake, B.W., C.P.L. Barkan & J.R. Edwards. 2011. Train Delay and Economic Impact of In-Service Failures of Railroad Rolling Stock. *Transportation Research Record: Journal of the Transportation Research Board* 2261: 124–133.
- Schmidt, S.R. & R.G. Launsby. 1994. *Understanding Industrial Designed Experiments*. Air Academy Press, Colorado Springs, CO.
- Sipilä, H. 2012. *Simulation of Rail Traffic: Applications with Timetable Construction and Delay Modeling*. Royal Institute of Technology, Department of Transport, Stockholm, Sweden. (Master's Thesis).
- Sogin, S.L., C.P.L. Barkan & M.R. Saat. 2011. Simulating the Effects of Higher Speed Passenger Trains in Single Track Freight Networks. In: *Proceedings of the 2011 Winter Simulation Conference*. IEEE, Phoenix, AZ, pp. 3679–3687.
- Sogin, S.L., C.P.L. Barkan, Y.-C. Lai & M.R. Saat. 2012a. Measuring the Impact of Additional Rail Traffic Using Highway & Railroad Metrics. In: *Proceedings of the 2012 Joint Rail Conference*. Philadelphia, PA.
- Sogin, S.L., C.P.L. Barkan, Y.-C. Lai & M.R. Saat. 2012b. Impact of Passenger Trains in Double Track Networks. In: *Proceedings of the 2012 Joint Rail Conference*. ASME, Philadelphia, PA.
- Sogin, S.L., Y.-C. Lai, C.T. Dick & C.P.L. Barkan. 2013a. Comparison of the Capacity of Single and Double Track Rail Lines Using Simulation Analyses. *Accepted, Transportation Research Record: Journal of the Transportation Research Board*.
- Sogin, S.L., Y.-C. Lai, C.T. Dick & C.P.L. Barkan. 2013b. Analyzing the Incremental Transition from Single to Double Track Railway Lines. In: *The 5th International Seminar on Railway Operations Modelling and Analysis*. Copenhagen, Denmark.
- Stathopoulos, A. & M. Karlaftis. 2001. Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record: Journal of the Transportation Research Board* 135–140.
- Transportation Research Board. 2010. *Highway Capacity Manual, 5th ed*. Transportation Research Board, Washington, D.C.



- Varian, H.R. 2006. *Intermediate Microeconomics: A Modern Approach*, 7th ed. W.W. Norton & Company.
- Vromans, M.J.C.M., R. Dekker & L.G. Kroon. 2006. Reliability and heterogeneity of railway services. *European Journal Of Operational Research* 172: 647–665.
- Washington, S.P., M.G. Karlaftis & F.L. Mannering. 2003. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL.
- White, T. 2006. Examination of Use of Delay as Standard Measurement of Railroad Capacity and Operation. In: *Proceedings of the 85th Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Williams, M.K. 2011. Using Simulation to Understand Bottlenecks, Delay Accumulation, and Rail Network Flow. In: *Proceedings of the Annual AREMA Conference*. AREMA, Minneapolis, MN.
- Wilson, E. 2012. *Rail Traffic Controller (RTC)*. No. 2.70 64K, Berkeley Simulation Software, Berkeley, CA.